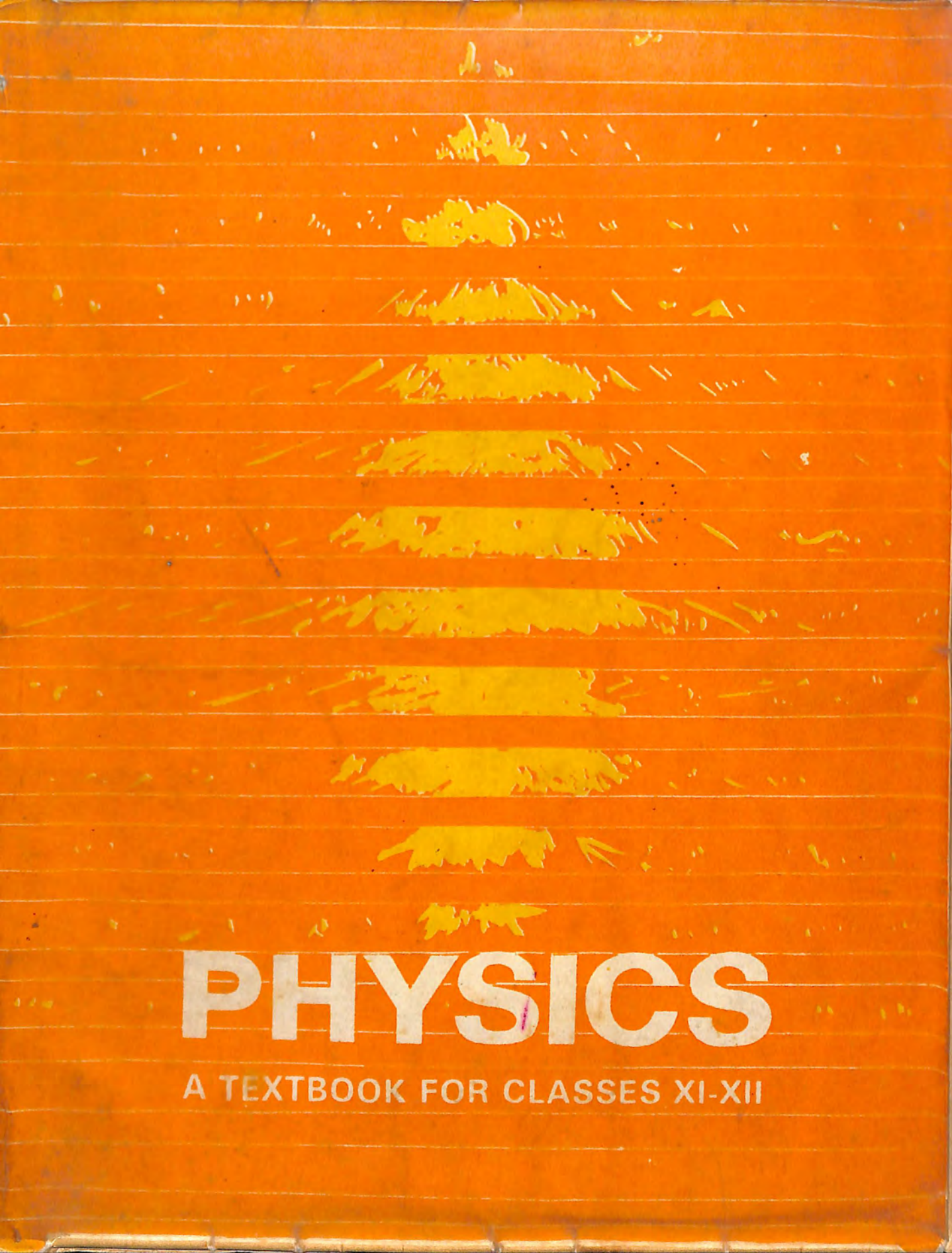
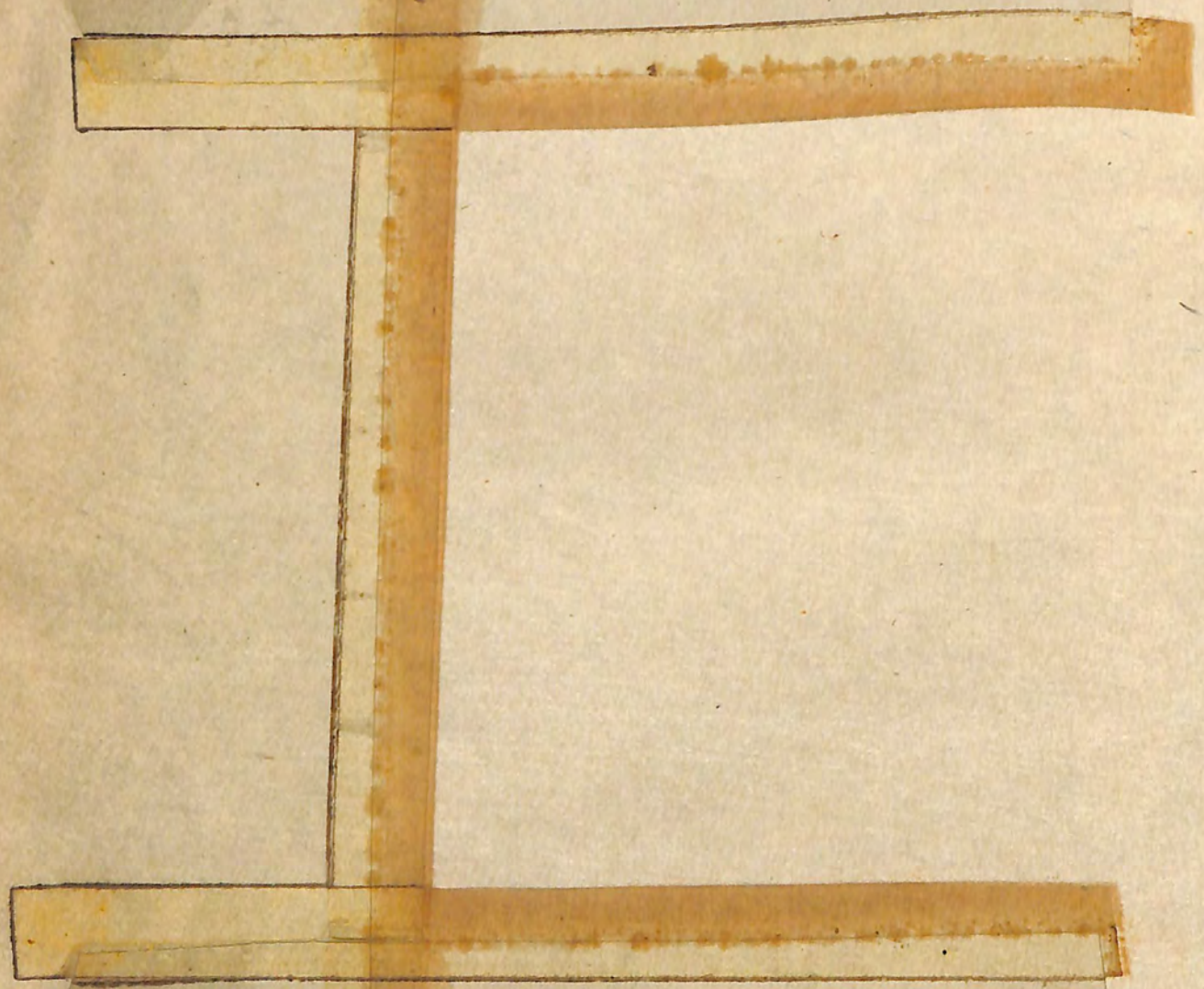
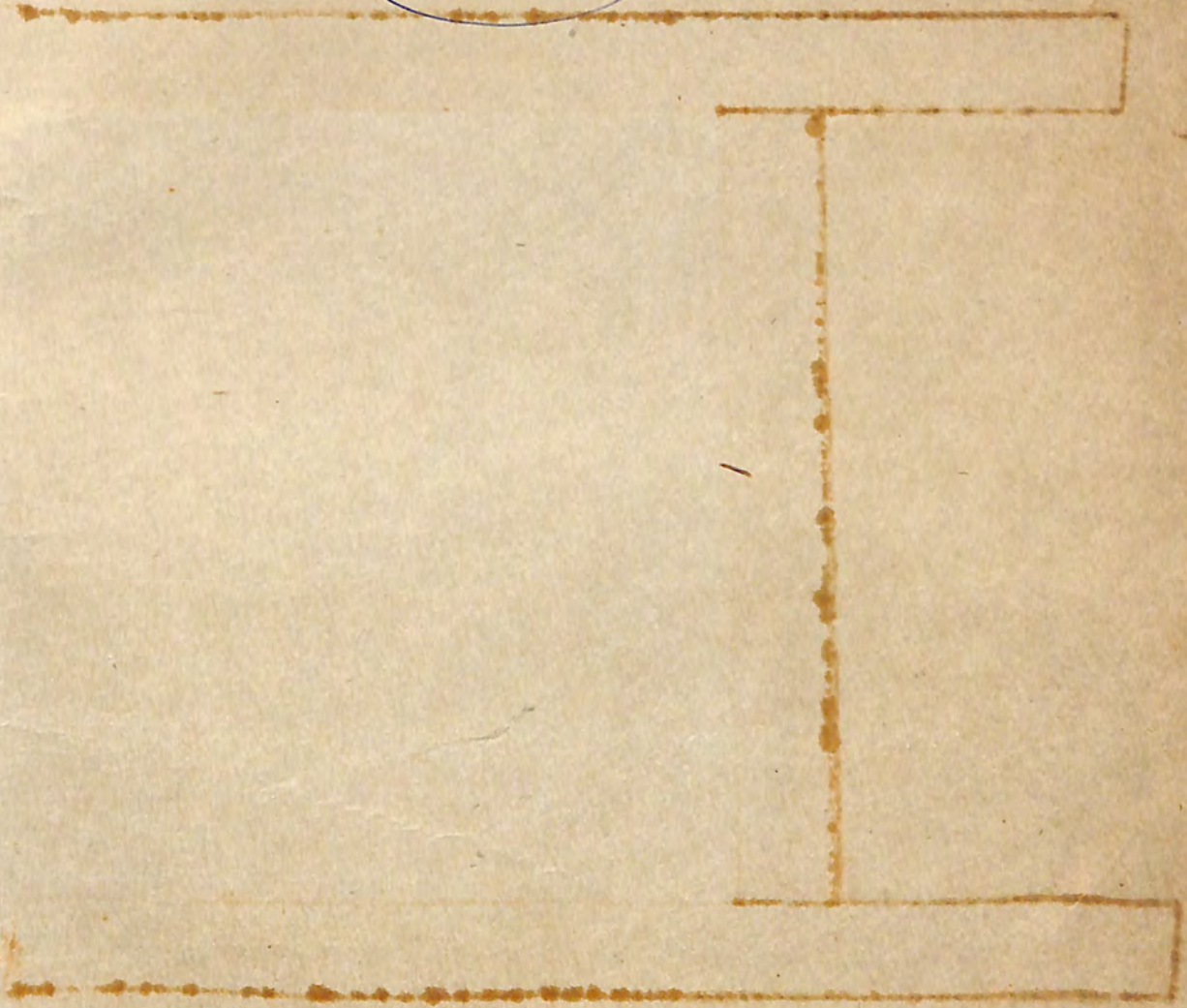


PHYSICS

A TEXTBOOK FOR CLASSES XI-XII



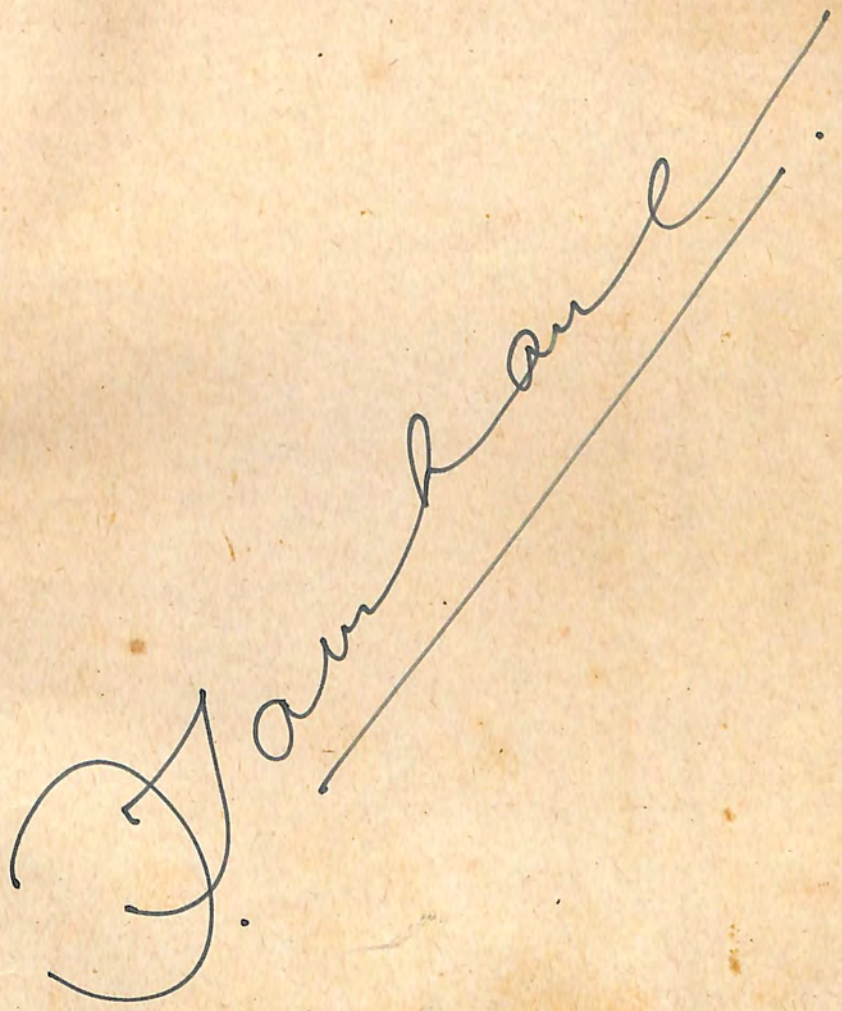
1752



EKLAVYA ... INDORE
CLASS No. 2100
ACC. No. 80405

TUSHAR TAMHANE.

16th. MAY-85



PHYSICS

A Textbook for Classes XI-XII

Editorial Board

Professor D.D. Pant (*Chairman*)
D.S.B. College
Kumaun University
Nainital

Professor Rais Ahmed
Vice-Chancellor
Kashmir University
Srinagar

Professor H. S. Hans
Punjab University
Chandigarh

Professor B. Sharan
Head, D.E.S.M.
NCERT
New Delhi

Professor B. Ramachandra Rao
Vice-Chairman
University Grants Commission
New Delhi

Professor M.S. Swami
Aligarh Muslim University
Aligarh

Professor S.K. Joshi
Roorkee University
Roorkee

Dr. S.G. Gangoli
Reader
NCERT
New Delhi

Authors

D.D. Pant
B. Ramachandra Rao
S.K. Joshi
H.S. Hans
K.D. Abhyankar
M.S. Swami
V.S. Murthy
U.S. Kushwaha
M.C. Durgapal
B. Sharan
S.G. Gangoli
D.P. Khandelwal
A.V. Patankar – Director, Western Regional Instrumentation Centre, BOMBAY.

Editors
D. D. Pant
S.K. Joshi

PHYSICS

A TEXTBOOK FOR CLASSES XI-XII

एकलव्य
अध्ययन केन्द्र बालाघेड़ी
होशंगाबाद फोन- 277261

Eklavya Library
Class No. 2133
Book No.
Shelf No. 64.5
Acc. No. 13373

विद्यया ऽमृतमश्नुते

एन सी ई आर टी
NCERT

राष्ट्रीय शैक्षिक अनुसंधान और प्रशिक्षण परिषद्
NATIONAL COUNCIL OF EDUCATIONAL RESEARCH AND TRAINING

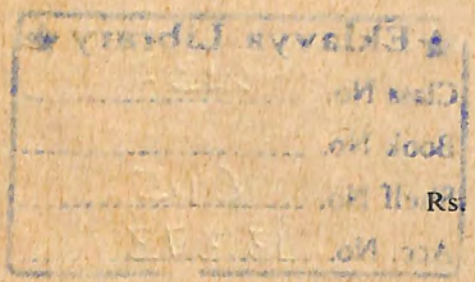
The book was brought out earlier in two parts. This is a combined edition.

First Edition (Combined)
July 1980
Asadha 1902

Reprinted
May 1981
Jyaishtha 1903
June 1982
Asadha 1904

P. D. 40T-RSS

© National Council of Educational Research and Training, 1977



Rs. 8.70

Published at the Publication Department, by V. K. Pandit, Secretary, National Council of Educational Research and Training, Sri Aurobindo Marg, New Delhi - 110 016 and printed at GPI Press, 19, T. H. Road, Madras-600 081.

Foreword

This is the revised version of the first edition of this book. The revision was undertaken on the basis of the feedback information which the Council received from its users. A notable feature of the revised edition is the appendix on mathematical notes. The appendix is expected to be covered by the teachers in the initial stages of the course. To make the book more functional some portions have been rewritten while some have been dropped.

This book is intended to meet the needs of the four semester course in Physics for classes XI-XII of the academic stream in the 10+2 pattern of education. Earlier it was brought out in different parts with a view to meeting urgent demand. Since the content is organised in units, the teacher has the flexibility of following any sequences of the units which he or she thinks would suit the requirements of the class.

We are grateful for the comments which we have received from the users of the earlier edition. These comments have helped us to revise the book. In particular we are appreciative of the efforts of those who participated in the workshop organised by the NCERT to evaluate and revise the book, and of the authors who willingly incorporated the suggestions in a very short time. Their names are listed separately.

Since detailed revision of both the parts of the book is under way suggestions for further improvement will be most welcome.

SHIB K. MITRA
Director
National Council of
Educational Research and Training

New Delhi
May, 1980

Acknowledgements

The National Council of Educational Research and Training is grateful to the following who participated in the Workshop to evaluate and revise the first edition of this textbook, held in NCERT from 30 March to April 1, 1978 :

Shri J.P. Aggarwal	..	Science Branch, Delhi Administration
Shri V.M. Arya	..	Science Branch, Delhi Administration
Shri R.S. Bharj	..	K.V.S., Andrews Ganj, New Delhi
Dr. R. Bhatia	..	N.C.E.R.T., New Delhi
Shri S. Chander	..	Delhi Public School, New Delhi
Dr. S.N. Dutta	..	N.C.E.R.T., New Delhi
Shri R.C. Dubey	..	Mothers International School, New Delhi
Shri L.C. Gupta	..	Sardar Patel Vidyalaya, New Delhi
Shri D.R. Jasuja	..	Summer Fields School, New Delhi
Shri R. Joshi	..	N.C.E.R.T., New Delhi
Shri G.S. Rautella	..	N.C.E.R.T., New Delhi
Shri M.M.S. Sajwan	..	N.C.E.R.T., New Delhi
Prof. B. Sharan	..	N.C.E.R.T., New Delhi
Shri D.P. Sharma	..	Bharatiya Vidya Bhawan, New Delhi
Shri C.M. Srivastava	..	Science Branch, Delhi Administration
Shri S.K. Srivastava	..	K.V.S., I.I.T., New Delhi
Dr. S.G. Gangoli (Co-ordintor)	..	N.C.E.R.T., New Delhi

Contents

PAGE

Foreword v

Unit 1 MEASUREMENT 1

1.1 Length and Time; 1.2 Measurement; 1.3 Systems of Units and SI Units; 1.4 Dimensions.

Unit 2 MOTION 9

2.1 Displacement; 2.2 Graphical Representation of Vectors; 2.3 Addition and Subtraction of Vectors; 2.4 Resolution of a Vector Along Two Given Directions; 2.5 Multiplication of a Vector by a Scalar; 2.6 Scalar Product of Two Vectors; 2.7 Vector Product of Two Vectors; 2.8 Kinematics; 2.9 Velocity; 2.10 Acceleration; 2.11 Equations of Motion; 2.12 Newton's First Law of Motion; 2.13 Newton's Second Law of Motion; 2.14 Principle of Conservation of Linear Momentum; 2.15 Newton's Third Law of Motion; 2.16 Inertial Mass; 2.17 Collision; 2.18 Impulse; 2.19 Motion in Two Dimensions—Projectile Motion; 2.20 The Range of a Projectile; 2.21 Work, Energy and Power; 2.22 Work; 2.23 Work Done by a Constant Force; 2.24 Work Done Against a Variable Force; 2.25 Energy; 2.26 Conservation of Energy; 2.27 Power; 2.28 Friction; 2.29 Origin of Friction; 2.30 Nature of Friction; 2.31 Laws of Friction; 2.32 Rolling Friction; 2.33 Lubrication;

Unit 3 CIRCULAR MOTION 41

3.1 Angular Velocity; 3.2 Uniform Circular Motion; 3.3 Banking of the Curve; 3.4 Planetary Motion and Kepler's Laws; 3.5 Kepler's Laws; 3.6 Newton's Universal Law of Gravitation; 3.7 Universal Gravitation Constant; 3.8 Gravitational Field; 3.9 Gravitational Potential Energy; 3.10 Earth's Satellites; 3.11 Escape Velocity; 3.12 Orbital Velocity; 3.13 Launching of a Satellite; 3.14 Rigid Body; 3.15 Angular Acceleration; 3.16 Rotational Kinetic Energy; 3.17 The Moment of Inertia of a Uniform Ring About an Axis Through its Centre O Perpendicular to the Plane of the Ring; 3.18 The Moment of a Uniform Circular Disc About an Axis Through its Centre Perpendicular to the Plane of the Disc; 3.19 Angular Momentum; 3.20 Torque; 3.21 Work Done by a Torque; 3.22 Torque as a Cross-product of Vectors R and F; 3.23 Analogy Between Linear Motion and Rotational Motion.

Unit 4 HARMONIC OSCILLATIONS 64

4.1 The Discription of a Harmonic Oscillation; 4.2 The Dynamics of a Harmonic Oscillation; 4.3 Some Examples of Harmonic Oscillations; 4.4 Particle Velocity and Acceleration; 4.5 Energy in Harmonic Oscillations; 4.6 Resonance.

Unit 5 WAVE MOTION 77

5.1 Waves on Water and Strings; 5.2 Sound Waves; 5.3 Velocities of Different Waves; 5.4 Harmonic Waves; 5.5 Graphical Representation of Wave Motion; 5.6 Phase and Phase Difference; 5.7 Wave Fronts; 5.8 Energy Transmission in Waves; 5.9 Reflection of Sound Waves; 5.10 Refraction of Sound Waves; 5.11 Polarisation in Waves; 5.12 The Doppler Effect.

Unit 6 SUPERPOSITION OF WAVES 97

6.1 Interference of Waves; 6.2 Diffraction of Waves; 6.3 Beats; 6.4 Stationary Waves; 6.5 Waves in Strings and Air Columns; 6.6 Acoustic Considerations in 7.7 Lasers; 7.8 Spectometer; 7.9 Various Kinds of Spectra

Unit 7 OPTICS 119

7.1 Nature of Light; 7.2 Interference of Light; 7.3 Coherent Sources; 7.4 Colours of Thin Films; 7.5 Laws of Reflection and Refraction; 7.6 Diffraction of Light; 7.7 Spectrometer; 7.8 Various Kinds of Spectra

Unit 8 KINETIC THEORY OF GASES 139

8.1 Introduction; 8.2 Assumptions for the Development of the Kinetic Theory of Gases; 8.3 Expression for the Pressure Exerted by a Gas; 8.4 Deduction of the Laws; 8.5 Relation Between Temperature and Kinetic Energy; 8.6 The Law of Equipartition of Energy; 8.7 Specific Heat of Gases (C_P and C_v).

Unit 9 PHYSICS OF THE ATOM 147

9.1 The Nature of Matter; 9.2 Cathode Rays; 9.3 Model of an Atom; 9.4 Bohr's Theory for the Hydrogen Atom; 9.5 Electron Configuration in Atom; 9.6 X-rays; 9.7 Photo Electric Effect; 9.8 Dual Nature of Radiation and Matter.

Unit 10 CONCEPTS OF SPACE, TIME, AND MASS IN RELATIVITY 171

10.1 Definition of Observer, Event and Frame of Reference; 10.2 The Principle of Relative Motion; 10.3 Galilean Transformations: 10.4 Newtonian Relativity Principle; 10.5 The Nature of Light; 10.6 The Michelson and Morley Experiment; 10.7 Special Theory of Relativity.

Unit 11 THERMODYNAMICS 179

11.1 Temperature; 11.2 Heat Energy; 11.3 Work; 11.4 Work and Heat; 11.5 Internal Energy; 11.6 First Law of Thermodynamics; 11.7 Applications of the First Law of Thermodynamics; 11.8 Conversion of Heat into Work—Heat Engine; 11.9 Second Law of Thermodynamics; 11.10

Second Law and Refrigerator; 11.11 Reversible Process; 11.12 Carnot Engine; 11.13 Efficiency of a Carnot Engine; 11.14 Reversibility of Carnot Engine; 11.15 Radiation; 11.16 Emission and Absorption of Radiation; 11.17 Black-body; 11.18 Kirchoff's Law; 11.19 Applications of Kirchoff's Law; 11.20 Stefan's Law; 11.21 Experimental Study of Black-body Radiation; 11.22 Wien's Displacements Law; 11.23 Pyrometers

Unit 12 LIQUIDS 197

12.1 Intermolecular Interactions; 12.2 Surface Tension; 12.3 Surface Energy; 12.4 Angle of Contact; 12.5 Pressure Difference across a Surface Film; 12.6 Capillarity; 12.7 Flow of Liquids; 12.8 Viscosity; 12.9 Bernoulli's Equation

Unit 13 ELECTRICITY 215

13.1 Electric Current; 13.2 Current Flow in a Metallic Conductor : Draft Velocity; 13.3 Heating Effects of Current : Joule's Law, 13.4 Variation of Resistance with Temperature; 13.5 Thermo-electric Effects; 13.6 Temperature Measurement by Thermocouples; 13.7 Magnetic Effects of Electric Current; 13.8 Biot-Savart Law—Magnetic Field due to Some Current Distributions; 13.9 Motion of a Charged Particle in a Uniform Magnetic Field; 13.10 Force on a Current carrying Conductor; 13.11 Torque on a Coil; 13.12 The Moving Coil Galvanometer.

Unit 14 ELECTROMAGNETIC INDUCTION 261

14.1 Magnetic Flux; 14.2 Laws of Electromagnetic Induction; 14.3 Methods of creating Induced E.M.F.; 14.4 The Generator or Dynamo; 14.5 Mutual Inductance; 14.6 Self-inductance; 14.7 Some Phenomena connected with Inductance; 14.8 Alternating Current; 14.9 A.C. Circuit containing Resistance only; 14.10 A.C. Circuit containing Inductance only; 14.11 A.C. Circuit containing Capacitance only; 14.12 LCR Circuit; 14.13 Transformer

Unit 15 UNIVERSE 237

15.1 Introduction; 15.2 Astronomical Instruments; 15.3 The Study of the Solar System; 15.4 Space Exploration and Search for Extra-terrestrial Life; 15.5 Study of the Sun; 15.6 The Stars; 15.7 Physical Properties of Stars; 15.8 Stellar Evolution; 15.9 The Milky Way; 15.10 Galaxies and the Universe

Unit 16 : Solids 285

16.1 Structure of Solids ; 16.2 Crystal Lattice ; 16.3 Amorphous Solids ; 16.4 Crystal Systems ; 16.5 Close Packed Structures ; 16.6 Forms of Binding in Solids ; 16.7 Electrons in Solids ; 16.8 Semi-conductors.

Unit 17 : Properties of Materials 302

17.1 Elastic Properties of Solids ; 17.2 Elastic Properties and Inter atomic Forces ; 17.3 Thermal Properties of Solids ; 17.4 Thermal Expansion ; 17.5 Thermal Conductivity ; 17.6 Magnetic Properties of Materials.

Unit 18 : Nuclear Physics 317

18.1 Atomic Nucleus ; 18.2 Isotopes ; 18.3 Mass Spectrometer ; 18.4 Nuclear Mass ; 18.5 Electron—Volt and Atomic mass Unit ; 18.6 Binding Energy ; 18.7 Nuclear Process ; 18.8 Nuclear Reactions ; 18.9 Properties of Neutrons ; 18.10 Artificial Radio-activity ; 18.11 Radio-isotopes ; 18.12 Use of Radioisotopes.

Unit 19 : Nuclear Inter actions 330

19.1 Fission ; 19.2 Energy released in fission ; 19.3 Chain Reactions and Controlled Chain Reaction ; 19.4 Nuclear Reactors ; 19.5 Fusion and Energy Generation in Sun and Stars ; 19.6 Radiation Hazards ; 19.7 Accelerators ; 19.8 Cosmic Rays ; 19.9 Elementary Particles.

Unit 20 : Electronic Devices 347

20.1 Essentials of an Electronic Device ; 20.2 Source of Electrons ; 20.3 Manipulation of Electrons ; 20.4 Diode ; 20.5 Semiconductor Diodes ; 20.6 Triodes ; 20.7 Transistor ; 20.8 Applications of Triodes and Transistors ; 20.9 Manipulation of Position : TV and CRO Tubes.

Appendix

A. Mathematical Notes 373

UNIT 1

Measurement

1.1 Length and Time

From early times, man became conscious of space and time. The cubit* and the foot were in use for measuring lengths and distances. The stars and the motion of the moon across the sky engaged the attention of inquiring minds. The study of astronomy started very early. The recurrence of the day and night and the seasons gave an idea of the passage of time. There is a clock built in the human system. The heart goes on beating from birth to death. The pulse was used for measuring short time intervals. From the positions of the sun and the heavenly bodies in the sky, man used to fix the time of the day and night, respectively.

We have learnt in earlier classes about the solar system, of which our planet earth is a member. The radius of the earth is 6.4×10^6m and its natural satellite, the moon, is at a distance of 3.85×10^8m from the earth. The mean distance of the sun from the earth is 1.50×10^{11}m. In the previous class, we have known about the Astronomical Unit (A.U.) of length. For astronomical distances, there is another larger unit called the *light-year* and it is equal to the distance light travels in one year. One light-year is approximately 9.47×10^{15}m.

The nearest star, outside the solar system, is Alpha Centauri and it is 4.3 light-years away from the earth. A *galaxy* is a group of stars. On a clear moonless night, we see in the sky a bright band of stars in the Milky Way. These are the stars of the galaxy to which the sun belongs. There are many billions** of stars in our galaxy. A typical large galaxy contains about 100 billion stars distributed over space 0.1 to 0.2 million light-years across (Fig. 1.1). The distance between one galaxy and a neighbouring galaxy varies from a few tenths to several million light-years.

Fig. 1.1 Spiral Galaxy in Andromeda. The two distinct white patches, one above and the other below, represent two other elliptical galaxies.

The galaxies are found to be receding from the earth at speeds increasing with their distances from the earth. Assuming that the

* Cubit is the length of the arm from the elbow to the tip of the middle finger.
** 1 billion = 10^9

limiting speed of these galaxies is the same as the speed of light, the theoretical limit of the size of the universe is 10 billion light-years.

At the other extreme, on the atomic and sub-atomic scale, the distances are very very minute, for example, the radius of a hydrogen atom is 5×10^{-11}m while the effective radius of a proton is 1.2×10^{-15}m.

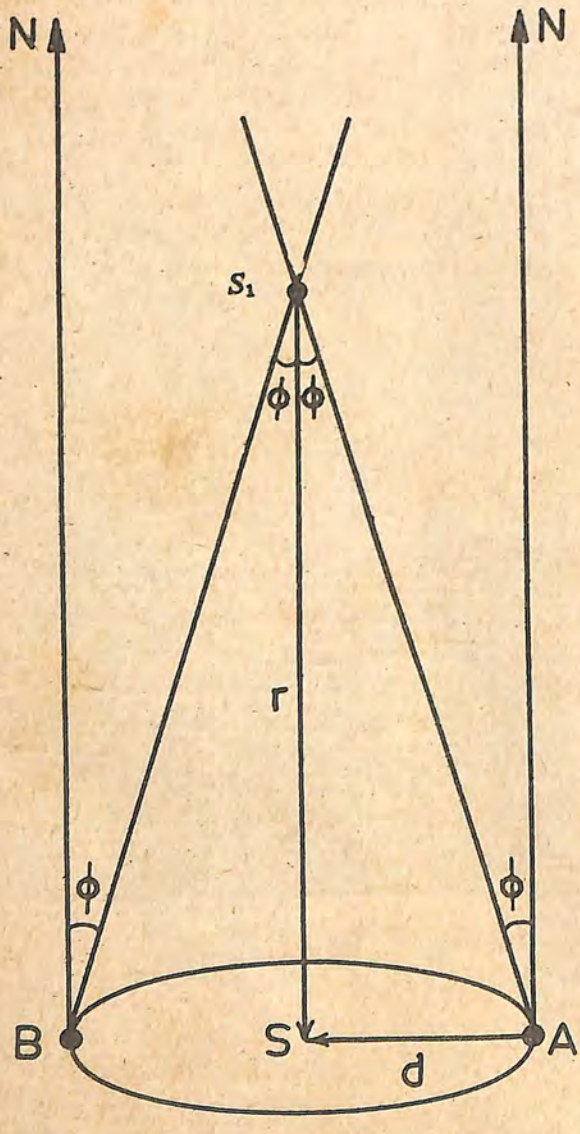


Fig. 1.2 Determination of the distance of a near-by star from the earth by the triangulation or parallax method. The directions BN, AN point towards the distant star N.

1.2 Measurement

Length

Suppose, we want to know the length of a room. Using a metre scale, we find out how many times the room is as long as the metre. If it is, say, 5 times the metre, the length of the room is 5 metres. This method of measurement by direct comparison cannot be applied in many cases. Indirect methods of measurement, therefore, become necessary.

For example, we are required to determine from a certain spot the distance of a hill not accessible for direct measurement. A gun may be fired from that spot directing it towards the hill and the time interval between the instant when the gun is fired and the instant when the echo is heard is measured. The speed of sound at that temperature is known from determinations made in the laboratory. The value of speed multiplied by half the time measured above gives the required distance. Even the indirect measurement of distances involves the use of the standard of length. In this case, the standard of length is used in the determination of the speed of sound. One assumption in finding the distance of the hill is that the value of the speed of sound determined in the subsidiary experiment holds good in the main determination also. Adopting the same principle, the distance of the moon from the earth was determined by sending radio waves from the earth to the moon.

Indirect methods only can be adopted for determining astronomical distances. The distances of the near-by stars can be determined by the triangulation or parallax method, using the diameter of the earth's orbit as the base-line.

To understand the principle, let us consider the simple case shown in Fig. 1.2. Let S_1 be the near-by star whose distance is to be determined. We choose, now, a far distant star (say, N) whose direction may be taken practically to be the same at all positions of the earth in its orbital motion.

Suppose, at some position of the earth (say, A) one measures, with the help of a telescope, the angle between the direction of the distant star N and the star S_1, i.e., the angle between AN and A S_1. Then, the observer waits for six months so that the earth is at B, a position diametrically opposite to the position A in the earth's orbit. At B, again, he measures the angle between the two directions. The sum of the magnitudes of these two angles is equal to the magnitude of the angle the star S_1 subtends on the earth's orbital diameter AB.

It can be easily seen with the particular example shown in Fig. 1.2 that

$$\frac{d}{r} = \tan \phi,$$

$$\text{so } r = \frac{d}{\tan \phi}. \qquad \ldots (1.1)$$

By the above method, only the distances of a few hundred stars which are relatively near to the earth can be determined. However, as the distance of a star increases, the angle ϕ decreases considerably and, thus, the above method cannot be relied upon. In the case of far-away stars, other indirect methods are used. For example, in one method, the intensity of the faint image of a distant star on a photographic plate is compared with the intensity of a near-by star which appears to be an equally strong source of light as the distant star. Applying the inverse square law (intensity proportional inversely to the square of the distance) and assuming the distance of the near-by star from the earth, the distance of the far-away star is determined. The two stars may not be exactly equally strong sources of light. Even though it is a rough method we can determine the order of magnitude of the distances of distant stars from the earth.

Hubble observed a shift in the spectral lines of the light from the distant galaxies. On the basis of the Doppler effect,* he explained that the distant galaxies are all receding away from us. Further studies indicated that the farther the galaxy is, the faster it is moving away. Hence, it is believed that the universe is expanding.

For measurement of very small distances such as those within molecules and atoms, only indirect methods can be adopted, for example, the determination of the size of the atom by Avogadro's hypothesis and the determination of the size of the gold nucleus by *alpha-particle* scattering experiments due to Rutherford. The diameter of the gold nucleus found by the α particle scattering experiment is of the order of 10^{-14}m.

The order of magnitude of lengths from astronomical to sub-atomic distances over which measurements can be made are given in Table 1.1.

TABLE 1.1

Order of Magnitude of Distances

Order of length in metres	Distance of that order
10^{26}	Distance to a distant quasar (quasi-stellar radio source)
10^{22}	Distance to the Great Nebula in Andromeda (nearest galaxy)
10^{20}	Diameter of our galaxy
10^{17}	Distance to the nearest star (Alpha Centauri)
10^{13}	Mean distance of the planet Pluto from the sun
10^{9}	Radius of the sun
10^{8}	Mean distance from the earth to the moon
10^{7}	Radius of the earth
10^{5}	Distance between Agra and Delhi
10^{2}	Length of the hockey field
10^{0}	Height of man
10^{-2}	Width of a finger
10^{-4}	Thickness of a sheet of paper
10^{-5}	Diameter of a red blood corpuscle
10^{-8}	Size of a virus
10^{-10}	Diameter of a hydrogen atom
10^{-14}	Size of a large nucleus
10^{-15}	Diameter of a proton

Time

For routine purposes, we like to know the time of the day so that we may keep to our work schedule. In scientific work, it is the measurement of time intervals that will be

* Refer to Art. 5.12

necessary. In early times, man used to judge the time of the day from the shadow cast by a pole. This principle is used in the sun-dial.

A phenomenon that repeats itself can be used as a measure of time, for example, the oscillations of a pendulum and a coiled spring. The measurement consists of counting the repetitions. The rotation of the earth on its axis is a repetitive phenomenon occurring in nature. This rotation determines the length of the day and was used as a standard of time from very early times. The day is subdivided into hours, minutes and seconds.

Table 1.2 gives the order of magnitude of time intervals that can be measured.

TABLE 1.2

Order of Magnitude of Time Intervals

Order of time intervals in seconds	An event of that order
10^{17}	Age of the earth
10^{15}	Time elapsed since the appearance of prehistorical animals
10^{13}	Time elapsed since man appeared on the earth
10^{11}	Age of the Pyramids of Gaza (Egypt)
10^{9}	Human life expectancy
10^{7}	Time for the earth to revolve once around the sun
10^{5}	Time for the earth to rotate once on its axis
10^{3}	Time taken by light from the sun to reach the earth
10^{2}	One minute
10^{0}	Time between heart-beats
10^{-2}	Time for an electric fan to make one complete rotation
10^{-4}	Time of the vibration of a high-frequency audible sound note
10^{-8}	Life time of an excited state of atom
10^{-11}	Time for light to pass through a glass-plate
10^{-15}	Time for an electron to revolve once around a proton in an atom of hydrogen
10^{-22}	Time for a single revolution of a proton in a nucleus

1.3 Systems of Units and SI Units

In physics, precision in measurement is very essential. Otherwise, the results of experiments based on measurements cannot be correlated. Hence, the basic set of units must be well defined. All mechanical quantities like density, velocity, momentum, force, and energy can be expressed in terms of three basic dimensional quantities, namely, length L, mass M and time T. In these, velocity which is displacement divided by time is written as

$$\text{velocity} = \frac{L}{T}.$$

A coherent system of units is a system based on a certain set of 'basic units' from which all derived units are obtained by multiplication or division without introducing numerical factors. The mks system is a coherent system of units for mechanics. m,k and s stand for the metre, the kilogram and the second which are the units for length, mass and time, respectively. In the International System of Units (abbreviated as SI), three additional dimensional quantities are used so as to cover all fields of physics. They are:

(i) In the field of electricity, the basic quantity is *current intensity*, the unit being the Ampere (symbol: A).

(ii) In the field of thermodynamics, the basic quantity is *thermodynamic temperature*, the unit being the degree-Kelvin (symbol: °K).

(iii) In the field of photometry, the basic quantity is *luminous intensity*, the unit being the candela (symbol: cd).

For the coherent system based on the six basic units mentioned above, the name International System of Units has been recommended by the International Conference on Weights and Measures in 1960. The units of this system are called SI-Units. In addition, there are 'dimensionless units', in particular the radian (symbol: rad) for a plane angle and the steradian (symbol: sr) for a solid angle. This system is favoured by scientists because of certain advantages over the other systems of units.

Length

The standard metre is the 'International prototype metre' preserved at the International Bureau of Weights and Measures at Sevres, near Paris. The international standard of length is a bar of platinum-iridium alloy. The distance between the two lines engraved on gold plugs near the ends of the bar, when the bar is at the temperature of melting ice, is called *one metre.* The speciality about this alloy is that it is least liable to be affected by time and environment.

The metre was originally intended to correspond to one ten-millionth of the earth's quadrant through Paris. Later on, more accurate determinations revealed that the standard metre is not exactly one millionth of the earth's quadrant. All the same, the original standard is continued. It may be noted that the unit chosen can be arbitrary. What is necessary is that it must be accepted internationally. Its value should not change with time. Scientists looked for a standard which is imperishable. In 1960, scientists agreed to adopt as a standard the wave length of a certain orange line in the spectrum of the Krypton isotope of atomic mass number 86, metre is defined in terms of this standard as exactly 1650763.73 wave lengths. They have now a standard which can be reproduced accurately whenever needed.

Multiples and sub-multiples of the metre are used for convenience, for example, kilometre (1000 m) and centimetre (1/100 m). Some of the commonly used metric prefixes are given in Table 1.3.

TABLE 1.3

Metric Prefixes

Prefix	Abbreviation	Meaning	Examples
tera	T	$\times 10^{12}$	
giga	G	$\times 10^{9}$	
mega	M	$\times 10^{6}$	1 megaton = 10^{6} tons
kilo	K	$\times 10^{3}$	1 kilometre = 10^{3}m
deci	d	$\times 10^{-1}$	
centi	c	$\times 10^{-2}$	1 centimetre = 10^{-2}m
milli	m	$\times 10^{-3}$	1 milliampere = 10^{-3} A
micro	μ	$\times 10^{-6}$	1 microvolt = 10^{-6}V
nano	n	$\times 10^{-9}$	1 nanosecond = 10^{-9} s
pico	p	$\times 10^{-12}$	1 picofarad = 10^{-12}F
femto	f	$\times 10^{-15}$	
atto	a	$\times 10^{-18}$	

NOTE: These prefixes can be used for multiples and sub-multiples of any unit, even though the unit is not in the metric system.

Mass

The standard kilogram* is the international prototype kilogram kept in the International Bureau of Weights and Measures at Sevres. It is a cylinder of nearly the same height and diameter and made of an alloy, 90 per cent platinum and 10 per cent iridium. This was adopted as the standard in 1889. The gram is 10^{-3}kg and the metric ton is 10^{3}kg. When the metric system was established in 1791 in France, the gram was intended to be the mass of one cubic centimetre of pure water at 4°C. More precise measurements have shown that this is not exactly true. And the prototype kilogram continues to be accepted as the standard. The pound mass and the yard which are the units of mass and length in the British system are now geared to the standard kilogram and metre, respectively, and are defined as 1 standard avoirdupois pound = 0.4535924277 kg and 1 standard yard = 0.9144 m.

The mass of two or more bodies can be compared by means of a common balance. We will learn in Unit 2 that mass is a quantitative measure of an attribute of matter called *inertia.* What we measure with a spring balance is not the mass of a body but the weight of the body which is the force of gravity upon it. Weight and mass should not be confused with each other.

Time

The day was defined as the interval of

* Copies of the standard kilogram and the standard metre are available at the National Physical Laboratory, New Delhi.

time between noon on one day and noon on the next day. This is the time taken by the earth for one complete rotation about its axis. The day is divided into 24 hours, each hour into 60 minutes and each minute into 60 seconds. So a second is 1/86400 of a day. Since the length of the day is found to vary during the course of the year, one (mean solar) second is defined to be 1/86400 of a mean solar day. The second is the unit of time. The unit of time is common to all systems of units. There are slight irregularities in the rotation of the earth on its axis. Also, the length of the year slowly changes though the variation is slight.

In 1956, the International Congress of Weights and Measures redefined the second, for scientific purposes requiring high precision, in terms of the earth's orbital motion about the sun. They defined the second to be the fraction 1/31556925.9747 of the tropical year 1900. There was search for a more reliable unit of time. Now the standard of time is based upon an 'atomic clock'. The present standard is based on the frequency of vibration of the cesium atom of atomic mass number 133. This is accurate to a few parts in 10^{11}. This was adopted in 1964 but was designated as 'temporary' because it is hoped that a standard based on some other atom such as hydrogen or thallium would be more reproducible.

It may be mentioned here that in the cgs system, the centimetre, the gram and the second are the three basic units of length, mass and time, respectively. In the fps system or the British system, the foot, the pound mass and the second are, respectively, the units of length, mass and time. In the scientific field, the SI units are now used.

1.4 Dimensions

All the derived units in mechanics can be expressed in terms of the fundamental units of length, mass and time. For example,

$$\text{Density} = \frac{\text{Mass}}{\text{Volume}} = \frac{\text{Mass}}{(\text{Length})^3} = \frac{(M)}{(L)^3}$$

$$= (ML^{-3})$$

$$\text{Velocity} = \frac{(L)}{(T)} = (LT^{-1})$$

$$\text{Acceleration} = \frac{\text{Change in Velocity}}{\text{Time}}$$

$$= \frac{(L)}{(T)(T)} = \frac{(L)}{(T^2)} = (LT^{-2}).$$

L, M, and T denote the dimensions of length, mass and time, respectively. The powers to which the fundamental units are raised to obtain the derived units are called the *dimensions of the derived units*. Volume has three dimensions in Length (L^3). The dimensions of density are one in mass and -3 in length and are written as ML^{-3}. Usually, the dimensions of quantities are expressed as equations as shown above. Such equations are called *dimensional equations*. Some quantities like angles and indices are dimensionless quantities. Numerical quantities have no dimensions. Quantities like specific gravity, which are expressed in the form of a ratio of two quantities having the same units, are dimensionless. The dimensions of a quantity will be the same whatever be the units in which it is measured.

By the use of dimensions we can check up the correctness of an equation. For example, consider the equation

$$v_f^2 = v_i^2 + 2as.$$

If the dimensions of each term in the equation are the same, then the equation is correct. v_f and v_i are the velocities, a is the acceleration and s is the distance travelled. The dimensions of v_f and v_i are the same, namely, $\frac{L}{T}$ and those of a are $\frac{L}{T^2}$. Now let us check up:

Dimensions of v_f^2 and v_i^2 are $\left|\frac{L}{T}\right|^2$

$= L^2 T^{-2}$.

Dimensions of $2as$ are $\frac{(L)}{(T^2)}(L)$

$= L^2 T^{-2}$.

Since the dimensions of each terms in the equation are the same, the equation is correct.

Dimensions can be used for deriving simple equations connecting physical quantities, provided the quantities taking part in the equation are known. For example, let us derive the expression for the periodic time of a simple pendulum.

In the case of a simple pendulum, the possible factors affecting the period of oscillation T are the length l of the pendulum, the mass m of the bob and the acceleration due to the gravity g at the place. Suppose the equation for the period of oscillation is

$$T = k\, l^x\, m^y\, g^z$$

where k is a dimensionless numerical constant. Substituting the dimensions of the quantities in the above equation, we have

$$(T) = (L)^x (M)^y (LT^{-2})^z$$

$$= (L)^{x+z} (M)^y (T)^{-2z}.$$

Equating the powers of L, M and T on both sides, we have

$$x + z = 0, y = 0 \text{ and } -2z = 1$$

$$\therefore x = \tfrac{1}{2}, y = 0 \text{ and } z = -\tfrac{1}{2}$$

$$\therefore T = k\, l^{\frac{1}{2}}\, m^0\, g^{-\frac{1}{2}} = k\sqrt{\frac{l}{g}}$$

On substituting the values in consistent units, the value of k can be obtained.

The dimensions for energy are the same as those for work. Table 1 gives the dimensions of some mechanical quantities.

TABLE 1

Dimensions

Quantity	How measured	Dimensions	Symbol
Length		L	m
Mass		M	kg
Time		T	s
Area	Length × Length	L^2	m^2
Volume	Length × Length × Length	L^3	m^3
Density	Mass/Volume	$L^{-3}M$	$kg\ m^{-3}$
Velocity	Distance/Time	LT^{-1}	ms^{-1}
Acceleration	Change in Velocity/Time	LT^{-2}	ms^{-2}
Momentum	Mass × Velocity	LMT^{-1}	$kg\ s^{-1}$
Force	Mass × Acceleration	LMT^{-2}	N
Work and Energy	Force × Distance	L^2MT^{-2}	J
Power	Work/Time	L^2MT^{-3}	W
Pressure	Force/Area	$L^{-1}MT^{-2}$	Pa
Angle	Arc/Radius	L^0	rad
Angular Velocity	Angle/Time	L^0T^{-1}	$radis^{-1}$
Angular Acceleration	Angular Velocity/Time	L^0T^{-2}	$radis^{-2}$
Moment of Inertia	Mass × (Distance)^2	L^2M	kgm^2
Torque	(Moment of Inertia) (Angular Acceleration)	$L^2\,MT^{-2}$	Nm
Frequency	No. of oscillations/time	T^{-1}	s^{-1}
Surface Tension	Force/Length	MT^{-2}	Nm^{-1}

EXERCISES

1.1 What is a coherent system of units?

1.2 What are the six basic units in the International System of Units?

1.3 What are the desirable characteristics of a physical standard?

1.4 In defining the standard of length we have to specify the temperature at which the measurement should be made. Are we justified in calling length a fundamental quantity if another physical quantity, temperature, has to be specified in choosing a standard?

1.5 What is the advantage in choosing the wave length of a light radiation as a standard of length?

1.6 Is the measure of an angle dependent upon the unit of length?

1.7 Suggest an indirect method of measuring the height of a tree on a sunny day.

1.8 Suggest a distance corresponding to each of the following order of length:

(i) 10^7m (ii) 10^4m (iii) 10^3m (iv) 10^2m (v) 10^{-3}m (vi) 10^{-6}m (vii) 10^{-14}m.

1.9 Mention some repetitive phenomena occurring in nature which could serve as time standards. What will be your criterion to select the most appropriate one as a time standard?

1.10 What is the advantage in expressing physical quantities in terms of dimensional equations. Mention two applications of dimensions.

1.11 Check the dimensional consistency in the case of the following equations:

(i) $S = v_o t + \frac{1}{2} at^2$ (ii) $v = v_o + at$ (iii) $FS = \frac{1}{2} mv^2 - \frac{1}{2} mv_o^2$

where S is the distance moved in time t, v_o and v are the velocities initially and at time t, a is acceleration, m is mass and F is the force acting.

1.12 The velocity of sound waves, v, through a medium may be assumed to depend on (i) the density of the medium, d, and (ii) its modulus of elasticity, E. Modulus of elasticity is a ratio of stress to strain and stress is force per unit area. Deduce by the method of dimensions the formula for the velocity of sound.

1.13 A liquid is flowing steadily through a pipe. Assume that the volume of the liquid issuing out per second depends on (i) the coefficient of viscosity of the liquid, η, (ii) the radius of the pipe, r and (iii) the pressure gradient along the pipe (pressure gradient is drop in pressure per unit length of pipe and it is equal to p/l where p is the pressure difference between the ends of the pipe and l is the length of the pipe).
The dimensions for the coefficient of viscosity are $ML^{-1}T^{-1}$. Deduce by the method of dimensions the formula for the volume of the liquid flowing out per second.

UNIT 2

Motion

2.1 Displacement

A change of position of a particle is called *displacement*. Let a particle move from position A to position B (Fig. 2.1). Its displacement is $\overrightarrow{AB}$, the arrow on top signifying that the particle has moved from A to B. If the particle moves from B to A, its displacement will be $\overrightarrow{BA}$. In either case, the magnitude of the displacement is the same but the direction in the second case is opposed to that in the first case. Hence,

$$\overrightarrow{AB} = -\overrightarrow{BA}. \qquad \ldots (2.1)$$

Fig. 2.1 A displacement vector.

A physical quantity that has a magnitude as well as a direction associated with it, is called a *vector*. Displacement is one such quantity. Therefore, displacement is a vector. Some examples of vectors are velocity, acceleration, momentum and force. The magnitude of a vector is called *modulus of the vector*. The modulus of the vector AB is denoted by $|\overrightarrow{AB}|$.

A quantity characterized only by its magnitude is called a *scalar*. Some examples of scalars are length, mass, time, energy and temperature.

2.2 Graphical Representation of Vectors

A vector is represented by an arrow. The length of the arrow is proportional to the magnitude of the vector and its orientation gives the direction of the vector. For example, a displacement of 4 km from the west to the east can be represented by an arrow PQ measuring 4 cm drawn along the west-east direction with the arrowhead pointing towards the east (Fig. 2.2). Analytically, a vector may be represented by

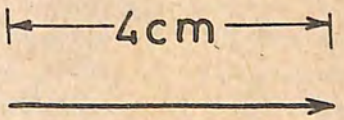


Fig. 2.2 Graphical representation of a vector.

a single letter with an arrow on top of the letter as $\vec{A}$ or by a bold face letter **A**. The latter can be used in print but, while writing, the single letter with an arrow on top will be more convenient. Ordinary *light face* type as well as *italics* are used to indicate the magnitudes of vectors as well as scalars.

Two vectors of the same physical quantity are equal if they are of the same magnitude and have the same direction. For example, the vectors **A** and **B** shown in Fig. 2.3 are equal. The vectors C and D

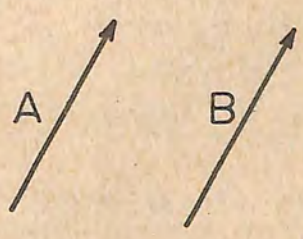


Fig. 2.3 Two equal vectors.

shown in Fig. 2.4 are equal in magnitude but opposite in direction.

Fig. 2.4 Vectors equal in magnitude, but opposite in direction.

$$\therefore \mathbf{C} = -\mathbf{D}$$

2.3 Addition and Subtraction of Vectors

Let a particle, initially at P, undergo a displacement of 4 m towards the east and then a displacement of 3 m towards the north. Let the final position be R (Fig. 2.5). Then, $\vec{PR}$ is the resultant displacement. (It may be

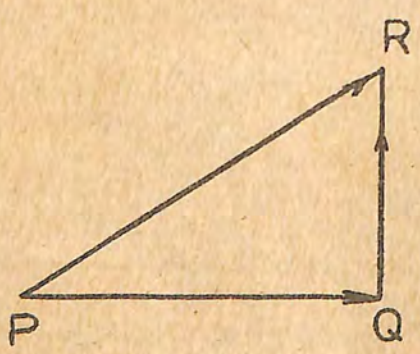


Fig. 2.5 Addition of two vectors: PQ = 4 m; QR = 3 m; PR = 5 m.

noted that its magnitude is only 5 m while the total distance moved is 7 m.) The above result can be written as

$$\vec{PQ} + \vec{QR} = \vec{PR}.$$

Likewise, the sum of three vectors A, B and C shown in Fig. 2.6a can be obtained by drawing the figure PQRS (Fig. 2.6b) in which the sides PQ, QR and RS represent the vectors A, B and C, respectively. PS represents the resultant D of the three given vectors.

$$\mathbf{A} + \mathbf{B} + \mathbf{C} = \mathbf{D}.$$

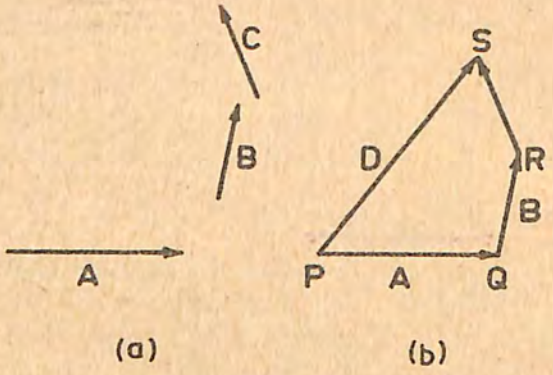


Fig. 2.6 Addition of three vectors.

In Fig. 2.7a, C = A + B. In Fig. 2.7b, the parallelogram PQRS is completed. In the triangle PSR, the side PS represents the

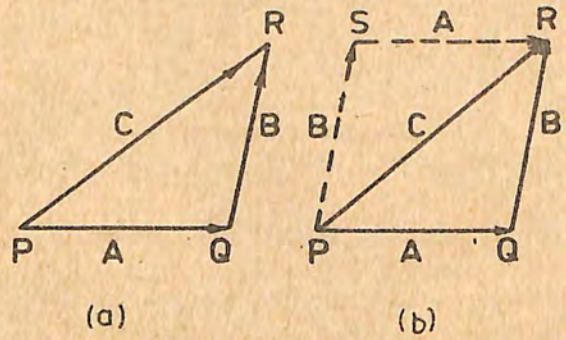


Fig. 2.7 Commutative law of vectors.

vector B and the side SR represents the vector A. PR represents the vector C.

$$\therefore \mathbf{B} + \mathbf{A} = \mathbf{C}, \text{ which is also equal to } \mathbf{A} + \mathbf{B}.$$

$$\therefore \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}. \qquad \ldots (2.2)$$

Thus, in adding two vectors A and B, the order of addition makes no difference. This is called *Commutative Law*.

Suppose, there are three vectors A, B and C. If we first add A to B, and then add C, we get the same result as that got by adding A to the sum (B + C). That is,

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \ldots (2.3)$$

This is called *Associative Law*.

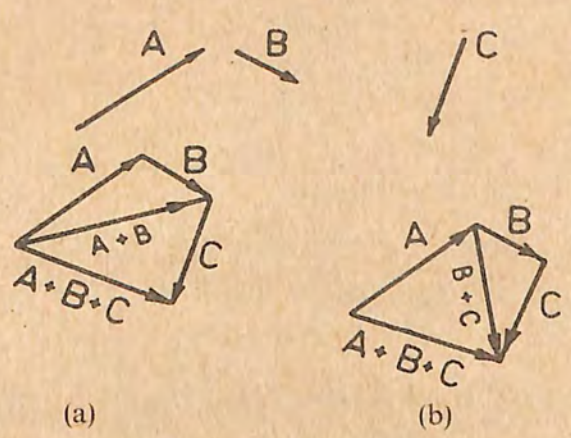


Fig. 2.8 Associative law of vectors.

Figs. 2.8a and 2.8b illustrate the associative principle of vectors.

EXAMPLE 2.1

A person has to cross a river by a boat from P to R exactly on the other side of the river (Fig. 2.9). In which direction should the person row the boat so as to finally arrive at R? Assume that he can row at 6 km/h in still water. The velocity of river is 3 km/hr.

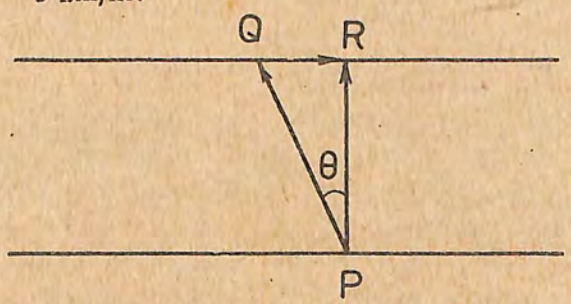


Fig. 2.9 A vector diagram showing motion in a river.

Let PQ be the direction in which the boat has to be rowed. The velocity of the boat along PQ, together with the velocity of the river, should yield a resultant velocity in the direction PR. If PQ represents the velocity 6 km/h and QR the velocity 3 km/h, then PR represents the resultant velocity of the boat. From the triangle PQR, we have

$$\text{Sin } \theta = \frac{QR}{PQ} = \frac{3}{6} = \frac{1}{2}.$$

$$\therefore \theta = 30°.$$

The boat should, therefore, be rowed upstream in the direction PQ, making an angle 30° with PR.

Let B be a given vector which has to be subtracted from another vector A Fig. 2.10a. In Fig. 2.10b, PQ represents A and QR represents —B. PR represents the sum of A and —B, that is, A — B.

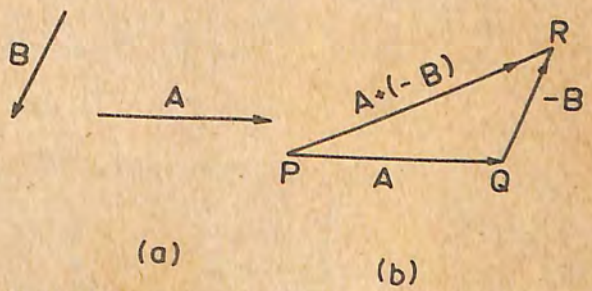


Fig. 2.10 Subtraction of one vector from another vector.

NOTE: It may be noted that addition and subtraction are possible only among vectors representing the same physical quantity.

2.4 Resolution of a Vector Along Two Given Directions

Let C be the given vector (Fig. 2.11) to be resolved along two mutually perpendicular directions X- and Y-. From the extrem-

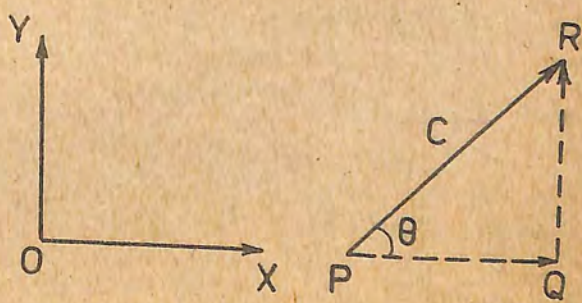


Fig. 2.11 Resolution of a vector along two mutually perpendicular directions.

ities P and R of the vector C, parallels are drawn, respectively, to the X-and Y-directions intersecting at Q. It is clear from the figure that PQ represents the component A along the X-direction and QR represents the component B along the Y-direction. If θ is the angle which C makes with the X-direction, the magnitudes of the components are given by

$$|A| = |C| \cos\theta \quad \text{and} \quad |B| = |C| \sin\theta.$$

By a similar geometric construction, the components of a given vector C along any two given directions can be graphically determined.

2.5 Multiplication of a Vector by a Scalar

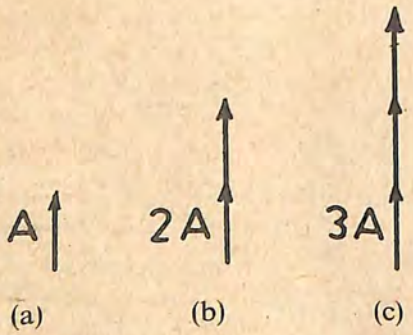


Fig. 2.12 Multiplication of a vector by a scalar.

When a vector A is multiplied by a scalar n, the resultant is nA. The direction of the resultant vector is the same as that of A but the magnitude is n times that of A. In Figs. 2.12b and 2.12c, the values of n are, respectively, 2 and 3.

2.6 Scalar Product of Two Vectors

The *scalar* (or *dot*) *product* of two vectors A and B is defined as the product of the magnitudes of the two vectors and the cosine of the angle between them. This can be written as

$$\mathbf{A}.\mathbf{B} = |A||B| \cos\theta. \qquad \ldots (2.4)$$

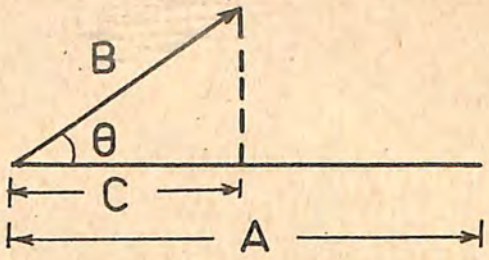


Fig. 2.13 Dot product of two vectors; **A.B** = AB cos θ; C = B cos θ.

$$\mathbf{A}.\mathbf{B} = AB\cos\theta; \quad C = B\cos\theta.$$

The dot product of two vectors is a scalar quantity. In the special case when θ is 90°, the dot product of two vectors becomes zero, since cos 90° is zero. When θ = 0, the dot product is the same as the product of the magnitudes of the vectors. The dot product can also be regarded as the product of the magnitudes of the vector A and projection (C in Fig. 2.13) of the vector B on A or *vice versa*.

The order of the factors in the case of a scalar product does not affect the product, since the factor cos θ remains the same in both the cases.

2.7 Vector Product of Two Vectors

Two vectors can be multiplied in such a way as to yield a vector. In this case, the product is referred to as the *vector* (or *cross*) *product*.

Let the vector product of two vectors A and B be equal to C. This is written as A × B = C and read as A cross B is equal to C.

If θ is the angle between the two vectors A and B, then the magnitude of the resulting vector C is defined by

$$|C| = |A||B| \sin\theta. \qquad \ldots (2.5)$$

The direction of C is perpendicular to the plane containing A and B in such a way that A, B and C form a right-handed coordinate system.

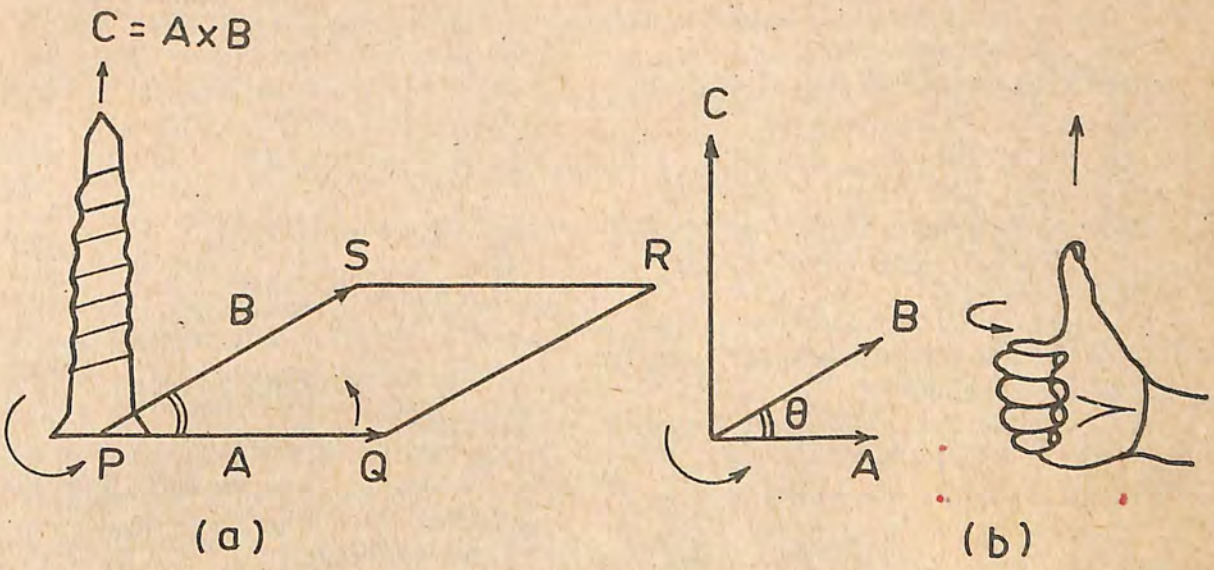


Fig. 2.14 Vector (or cross) product of two vectors.

The direction of the vector C can be conveniently obtained in the following way. Let a right-handed screw (Fig. 2.14a) whose axis is perpendicular to the plane containing the vectors A and B be turned from A to B through the angle θ between them as if driving the screw in. The direction along which the screw advances gives the direction of the resultant vector C. In the figure, this is directed upwards.

There is another way of representing the direction. Let the right hand be held as shown in Fig. 2.14b, with the thumb erect and the fingers folded round. Let the direction of the thumb be perpendicular to the plane containing the vectors A and B. If the direction of rotation of the vector from A to B is the same as the direction of folding of the fingers, then the thumb points towards the direction of the resultant C of the vector product. The magnitude of this vector may be regarded as equal to the area of the parallelogram PQRS formed with A and B as adjacent sides, as indicated in Fig. 2.14a.

B × A *the same as* A × B?

In this case, applying the right-handed screw rule, the rotation of the right-handed screw should be from B to A through the angle θ as if driving the screw in. Then, the screw advances downwards, as shown in Fig. 2.15. The direction of C is perpendicular to the plane containing the vectors A and B in such a way that they form a left-handed co-ordinate system.

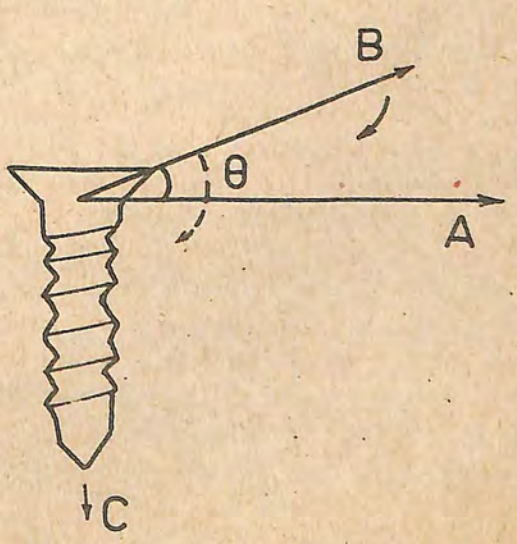


Fig. 2.15 Vector product changes the sign when the order of vectors in the product is reversed: **A** × **B** = −**B** × **A**.

Hence, it is clear that the vector product changes the sign when the order of the vectors in the product is reversed. Therefore,

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A}. \qquad \ldots (2.6)$$

This is in contrast with the scalar product of two vectors which is not affected by the order of the factors.

Force on a moving charge in a magnetic field is an example of a physical quantity that is a vector product.

2.8 Kinematics

Kinematics deals with the study of motion without taking into account the factors responsible for motion. In everyday life, we generally come across three types of motion: linear motion in which a body moves along a line (for example, a car going on a road), rotational motion wherein a body rotates about a fixed axis (for example, the wheel of a cart or the blades of an electric fan rotating about the axis of the shaft), and vibratory or to-and-fro oscillatory motion of the type we have in a swing or the pendulum of a wall clock. In this section, we shall consider the simple case of the linear motion of a body, treating it as a particle.

2.9 Velocity

Velocity is the Rate of Displacement

Since displacement is a vector, velocity is also a vector. It has both magnitude and direction. The dimensions of velocity are LT^{-1}.

To find the velocity of a particle at any instant of time, we should consider a very short time interval Δt such that the instant chosen corresponds to the middle of the interval t and find out the displacement Δs during this short time of interval. Then, $\frac{\Delta s}{\Delta t}$ gives the velocity of the particle at that instant of time. The shorter the length of the time interval, the more accurately will this ratio $\frac{\Delta s}{\Delta t}$ give the instantaneous velocity of the particle.

We may, therefore, write:

$$\text{Instantaneous velocity } v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \frac{ds}{dt} \qquad \ldots (2.7)$$

where $\frac{ds}{dt}$ is the derivative of displacement s with respect to time t.

The velocity of a particle is said to be *uniform* if the displacement of the particle in equal intervals of time are equal, however small these intervals may be.

Suppose, a particle is moving with a uniform velocity of 60 km/h in a particular direction, say, towards the east. It implies that the direction of motion of the particle is the same throughout, namely, towards the east and that in every half-hour, it moves 30 km; in every minute, one km; in every second, one sixtieth of a km; and in every one hundredth of a second, 1/6000 of a km. Even if we consider still shorter equal intervals of time, the distances travelled in those intervals will be equal. This is the meaning of uniform velocity.

In the case of a particle moving with variable velocity, the total distance covered divided by the time gives the average speed of the particle. Consider a particle moving from the position A to the position B in a time t following an irregular path as in

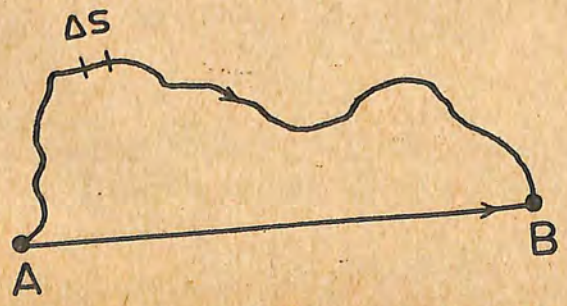


Fig. 2.16 A particle moving with the variable velocity.

Fig. 2.16. It can be shown that the average velocity v_{av}^{*} of the particle is given by the vector distance AB divided by the time t.

In this section, we shall limit our study only to motion in one dimension, i.e., along a straight line.

2.10 Acceleration

A body has acceleration when its velocity changes with time. The change in velocity consists of a change in the magnitude or change in the direction of motion or both. *The rate of change of velocity with time is called acceleration.*

If Δv is the change in velocity of a particle in a short time interval Δt, then the acceleration a is given by

$$a = \frac{\Delta v}{\Delta t}$$

where a is a vector since Δv is a vector.

If the acceleration is variable, the instantaneous acceleration at time t is given by

$$a = \lim_{t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} \qquad \ldots (2.8)$$

If the velocity changes uniformly with time, the acceleration is uniform and $\frac{dv}{dt}$ is constant. The dimensions of the acceleration are LT^{-2}.

Negative acceleration is called *retardation*. When the speed of a car is increasing, we say that the car is accelerating. When its speed is decreasing, it is subject to a retardation and we say that the car is decelerating.

2.11 Equations of Motion

Let a particle be in linear motion with constant acceleration a. Then

$$\frac{dv}{dt} = a$$

or $dv = a\,dt.$

Integrating both sides,

$$\int dv = \int a\,dt$$

or $v = at + k.$

where k is the constant of integration.

If v_i and v_f are the initial and final velocities of a particle at the beginning and end of a time interval t, we have

$v = v_i$ when $t = 0$

so that $k = v_i$.

$$\therefore \quad v_f = v_i + at. \qquad \ldots (2.9)$$

The velocity v at any instant of time t from the start is given by

$$v = v_i + at.$$

The distance Δs covered by the particle in an infinitesimally short time interval Δt over which time the velocity v can be assumed to be constant is given by

$$s = v\,dt = (v_i + at)\,dt.$$

Integrating, we get

$$\int ds = \int (v_i + at)\,dt.$$

$$\therefore \quad s = v_i t + \frac{1}{2}at^2 + k'.$$

since $s = 0$ when $t = 0$, k' is equal to 0

$$s = v_i t + \frac{1}{2}at^2. \qquad \ldots (2.10)$$

---

* From the definition of average, v_{av} can be written as

$$v_{av} = \frac{\Sigma v\,\Delta t}{\Sigma \Delta t} = \frac{\int v dt}{\int dt} = \frac{1}{t}\int v\,dt$$

Substituting for v from the equation (2.7), we have

$$v_{av} = \frac{1}{t}\int \frac{ds}{dt}\,dt$$

$$= \frac{1}{t}\int ds = \frac{\vec{AB}}{t}.$$

By eliminating t between the equations (2.9) and (2.10), we get

$$v_f^2 = v_i^2 + 2as. \quad \ldots (2.11)$$

The three equations (2.9), (2.10) and (2.11) are the three possible equations of motion between the three variable quantities, namely, v, s and t.

In the case of a freely falling body, the acceleration is due to gravity and is denoted by g which varies from place to place on the earth.

EXAMPLE 2.2

An automobile moving due north at the rate of 72 km/h^{-1} is brought to rest in 4 seconds by the application of brakes. Find the value of the acceleration, assuming it to be constant. What is the distance moved by the automobile from the instant the brakes are applied before it comes to rest?

The initial velocity of the automobile v_i

$$= 72 \text{ km/h}^{-1} = \frac{72 \times 1000}{60 \times 60}$$
$$= 20 \text{ ms}^{-1}$$

Since it is brought to rest, the final velocity $v_f = 0$

Let the acceleration during this period be a.

Substituting the values for v_f, v_i and t in the equation (2.9), we get

$$0 = 20 + 4a.$$
$$\therefore a = -5 \text{ ms}^{-2}.$$

The ratardation of the car is 5 ms^{-2}.

Substituting the values for v_f, v_i and a in the equation (2.11), we have

$$0 - (20)^2 = 2(-5) \times s$$
or $s = 40$ m.

After applying the brakes, the car comes to rest in 40 metres.

EXAMPLE 2.3

From the top of a multi-storeyed building, 39.2 metres tall, a boy projects a stone vertically upwards with an initial velocity of 9.8 ms^{-1} such that it finally drops to the ground.

(i) When will the stone reach the ground?
(ii) When will it pass through the point of projection?
(iii) What will be its velocity before striking the ground? (Assume the value of g to be 9.8 ms^{-2}.)

Due regard should be paid to the direction of each quantity in solving these problems. The +ve sign may be given to the quantities in the upward direction and the −ve sign to those in the downward direction.

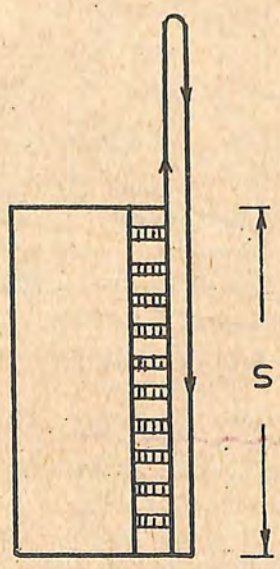


Fig. 2.17 A stone projected vertically upwards from the top of a multi-storeyed building. The arrow indicates the path of the stone.

(i) By the time the stone reaches the ground, the net distance travelled by

the stone is 39.2 m in the downward direction.

$$\therefore \quad s = -39.2\ m.$$

$$g = -9.8\ ms^{-2}.$$

$$v_i = 9.8\ ms^{-1}.$$

Substituting the values for s, g and v_i in the equation (2.10), we have

$$-39.2 = 9.8\,t - \frac{1}{2} \times 9.8\,t^2$$

$$\text{or } (t-4)(t+2) = 0.$$

$$\therefore \quad t = 4 \text{ or } -2.$$

The −ve value of t is not admissible. The stone reaches the ground 4 seconds after it is projected.

(ii) When the stone passes through the point of projection, the net displacement is zero.

∴ We have from the equation (2.10):

$$0 = 9.8\,t - \frac{1}{2} \times 9.8\,t^2$$

$$\text{or } (t-2)\,t = 0.$$

$$\therefore \quad t = 0 \text{ or } 2.$$

$t = 0$ corresponds to the start of the motion.

The stone passes through the point of projection 2 seconds after it is projected.

(iii) Let v_f be the velocity of the stone just before it strikes the ground. From the equation (2.9),

$$\text{we have } v_f = 9.8 - 9.8 \times 4$$

$$\therefore \quad v_f = -29.4\ ms^{-1}$$

The −ve sign indicates that the velocity is downward.

EXAMPLE 2.4

A food packet is released from a helicopter which is rising steadily at 2 ms^{-1}. After 2 seconds,

(i) what is the velocity of the packet?
(ii) how far is it below the helicopter?
Given $g = 9.8\ ms^{-2}$.

In this example $v_i = 2$ m/s, $g = -9.8$ m/s^2, $t = 2$ s.

(i) Substituting these values in the equation (2.9), we get

$$v_f = 2 - 2 \times 9.8 = -17.6\ ms^{-1}.$$

The velocity of the packet after 2 s is 17.6 ms^{-1} downwards.

(ii) Substituting the values for v_i, v_f and g in the equation (2.11), we have

$$(-17.6)^2 = (2)^2 - 2 \times 9.8 \times s$$

from which

$$s = -15.6\ m.$$

In 2 s, the helicopter has risen 4 m above the point at which the packet is dropped.

∴ The packet is 19.6m below the helicopter.

2.12 Newton's First Law of Motion

It is common experience that a body at rest continues to remain at rest unless it is disturbed by some external agency.

Let us consider bodies in motion. The Greek philosopher Aristotle (B.C. 384–322) stated that a constant force should continuously act on a body to keep it moving with constant velocity. No one opposed his statement for nearly two thousand years. Galileo (1564–1642) was the first to infer from his study of motion on the surface of an inclined plane that when a velocity is imparted to a body it continues to move with the same velocity provided no force acts on it. Such a

condition is realised in the case of a body moving on a horizontal plane when the surfaces are perfectly smooth. He observed that, when the surfaces were smooth, bodies moved for longer time and with fairly constant velocity on a horizontal plane. He concluded that, in the ideal situation, motion along a horizontal plane should be perpetual.

In actual practice, it is hard to find perfectly smooth surfaces offering no frictional resistance at all to motion.

By means of the 'Linear Air Track' (Fig. 2.18), an almost frictionless track can be secured. This consists of a long hollow tube of aluminium, a section of which is as shown in Fig. 2.18a. A number of small holes are drilled on the top of the tube. Air under pressure (from an air compressor) is forced into one end of the tube. The air comes out through the small holes and forms a cushion of air under a cart placed on top of the track (Fig. 2.18b). Given a push, the cart moves almost floating on the cushion of air. Thus, friction between the surfaces in contact is almost eliminated.

Keeping the air track horizontal*, let the cart be given a push. Then, as the cart moves, flash photographs** are taken at equal short intervals of time. From the pictures obtained, the distances covered in these equal time intervals are estimated. They will be found to be equal throughout the motion showing that the cart moved with almost constant velocity. Thus, when all external forces, including friction, are eliminated, the object in motion continues to be in uniform motion along a straight line. We may sum up by saying that every object has the tendency to continue in its own state, namely, the state of uniform motion in a straight line unless acted on by an external agency. This property of bodies is called *inertia*. Galileo was the first to recognize this property. Galileo's law of inertia implies that when no force acts on an object, it remains at rest if it is already at rest or moves along a straight path with constant speed, if it is already in uniform motion along a straight line.

After Galileo, it was Newton (1642–1727) who made a systematic study of motion and extended the ideas of Galileo. He arrived at three laws of motion which go by his name.

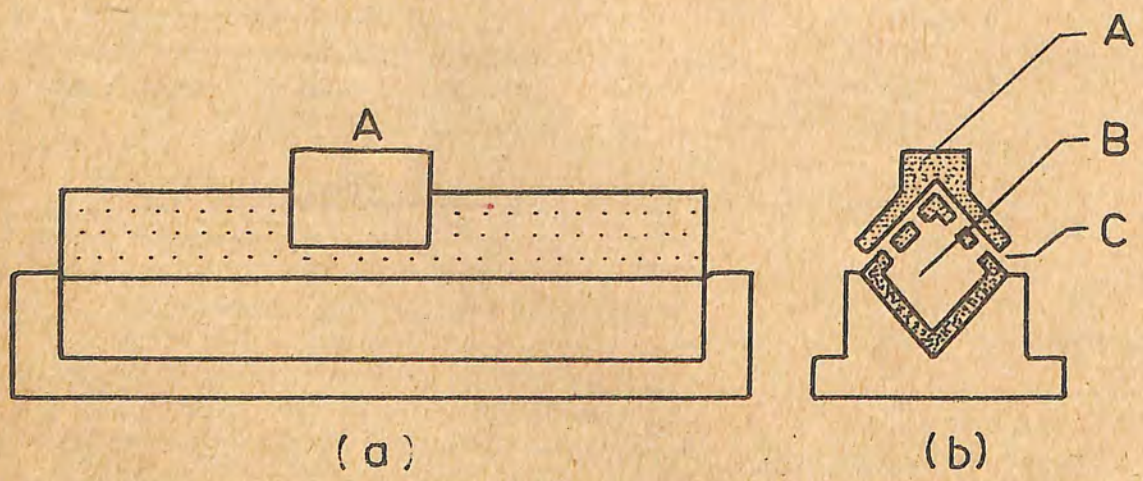


Fig. 2.18 Linear air track apparatus: A—cart; B—air under pressure; C—film of air.

* By keeping the air track horizontal, the force of gravity will be eliminated.

** A sequence of pictures, at regular intervals of time, can be taken on a single piece of film by multi-flash photography in the following way. The camera shutter is opened in the dark and the object to be photographed is illuminated by a rapidly flashing light which may be from an electronic flash tube. When the event is completed, the camera shutter is closed.

The first law of motion is as follows:

> Every body persists in its state of rest or of uniform motion in a straight line unless it is compelled to change that state by a force impressed on it. This is in fact a re-statement of Galileo's law of inertia.

We have many examples of inertia in every day life. It is due to inertia that passengers experience a jerk in the forward direction when a bus in motion is suddenly stopped. The standing passengers may even fall down. While their feet are suddenly brought to rest, the remaining part of the body continues to move forward. It is for the same reason that a person who gets down from a running train has a tendency to fall in the direction of motion of the train.

2.13 Newton's Second Law of Motion

A particle in motion has momentum. It is the product of its mass and velocity. If m is the mass of the particle and v its velocity, then its linear momentum p is defined as:

$$p = mv \qquad \ldots (2.12)$$

Since velocity v is a vector, the linear momentum p is also a vector and its direction is the same as that of v.

When a force acts on a body, its velocity changes. Consequently, its momentum also changes. Let us suppose that the same force acts on different bodies for the same length of time. The change in velocity produced in a body of large mass will be less than that produced in the one with a smaller mass. On the other hand, if we consider the change in momentum produced, it will be the same in each case.

Newton stated his second law in terms of the rate of change of momentum as follows:

> *The rate of change of momentum of a body is proportional to the force acting on the body and is in the direction of the force.*

The momentum p of a body of mass m moving with velocity v is already defined in the equation (2.12)

Differentiating this equation with respect to time t, we get

$$\frac{dp}{dt} = \frac{d(mv)}{dt}, \text{ or } \frac{dp}{dt} = m\frac{dv}{dt},$$

taking m as a constant

$= ma$, where a is the acceleration produced.

According to Newton's second law, stated above, the rate of change of momentum $\frac{dp}{dt}$ is proportional to the force F, but since $\frac{dp}{dt} = ma$ we have

$$F \propto ma \text{ or } F = kma. \qquad \ldots (2.13)$$

where k is a constant.

A unit force can be suitably defined so as to make the value of k equal to one. A unit force is that which acting on unit mass produces in it unit acceleration in the direction of the force.

In the MKS system, a unit is that which acting on a body of mass 1 kg produces in it acceleration of 1 ms^{-2}. This unit is called a newton (symbol N).

In the CGS system, unit force is that force which acting on a body of mass 1 g produces in it acceleration of 1 cm s^{-2}. This is called a *dyne*.

The dimensions of force are MLT^{-2}.

2.14 Principle of Conservation of Linear Momentum

The principle of conservation of linear momentum may be stated as follows:

> While the external force acting on a system of several particles is zero, the total linear momentum of the system is conserved (constant). The total linear momentum is the vector sum of the linear momentum of each particle in the system.

This law of conservation of linear momentum is a fundamental law, like the law of conservation of energy. Both these laws are of great importance in physics.

Suppose, there is only one object in an isolated system (with no external forces acting on it). If the object is moving with a velocity v, according to the first law of motion, it should continue to move with the same velocity; hence, its linear momentum mv remains constant.

Let us there be two objects in an isolated system and let them be interacting with each other (e.g., by colliding). If we consider any one object, its velocity will be changing due to the interaction; hence, its linear momentum will be changing. The same will be the case with the second body also. Since there are no external forces acting on the system, the vector sum of the linear momenta of the two objects would remain the same.

The same thing applies even if there are three or more objects interacting among themselves in an isolated system. This is the meaning of the principle of conservation of linear momentum.

A formal proof of this is as follows:

Suppose, in a system, there are *n* particles of mass $m_1, m_2, \ldots m_n$ with velocities $V_1, V_2, \ldots V_n$, respectively. Then the total linear momentum P is given by the vector sum of the linear momenta of all the particles in the system. That is

$$P = m_1 V_1 + m_2 V_2 + m_3 V_3 + \ldots m_n V_n.$$

But, we have

$$MV_{c.m} = m_1 V_1 + m_2 V_2 + m_3 V_3 \ldots m_n V_n.$$

where M is the total mass of all the particles in the system and $V_{c.m}$ is the velocity of the 'centre of mass'* of the system. That is, the total linear momentum of a system of particles is the product of the total mass of the system and the velocity of its centre of mass.

From the above equations, we have

$$P = M V_{c.m.} \qquad \ldots (2.14)$$

Differentiating with respect to time, we have

$$\frac{dP}{dt} = M \frac{dV_{c.m}}{dt} = Ma_{c.m}$$

where a_{cm} is the acceleration of the centre of mass of the system. According to Newton's second law of motion, we have

$$\frac{dP}{dt} = \text{resultant external force on the system} = F_{ext}.$$

When the resultant external force on the system is zero, we have $F_{ext} = \frac{dP}{dt} = 0$, in which case P = Constant, i.e., the total linear momentum (vector sum) of the system is constant. This is the principle of Conservation of linear momentum which is a direct consequence of Newton's Second Law of Motion.

2.15 Newton's Third Law of Motion

Let m_1 and m_2 be the mass, of two bodies constituting an isolated system. Suppose,

---

* The centre of mass of a system of particles is that single point within the system which moves in the same way in which a single particle having the total mass of the system and subject to the same external forces would move. The centre of mass of a system of particles depends on the mass of the particles and the positions of the particles relative to one another.

If there are *n* particles of mass $m_1, m_2, m_3, \ldots m_n$ distributed in space and if the co-ordinates of these particles with respect to three co-ordinate axes, x, y, and z are $(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3), \ldots (x_n, y_n, z_n)$, respectively, then the co-ordinates $(x_{c.m}, y_{c.m}, z_{c.m})$ of the centre of mass of the system of particles are given by

$$x_{c.m} = \frac{m_1 x_1 + m_2 x_2 + \ldots + m_n x_n}{m_1 + m_2 + \ldots + m_n} = \frac{\sum_1^n m_i y_i}{M}$$

Similarly, we have $y_{c.m} = \frac{\sum_1^n m_i x_i}{M}$ and $z_{c.m} = \frac{\sum_1^n m_i z_i}{M}$

they are moving along the same line. Let one body interact with the other body. The velocities of the bodies change, consequently there will be a change in momentum. Let Δp_1 and Δp_2 denote the change in momentum of m_1 and m_2, respectively, in time Δt. According to the law of conservation of linear momentum, the net change in the linear momentum in an isolated system must be zero.

$$\therefore \Delta p_1 + \Delta p_2 = 0$$

$$\text{or} \quad \Delta p_2 = -\Delta p_1. \qquad \ldots (2.15)$$

Dividing by the time Δt and taking the limiting case,

$$\frac{dp_2}{dt} = -\frac{dp_1}{dt}$$

i.e., Rate of change of momentum of m_2 = − Rate of change of momentum of m_1

Force on m_2 = − Force on m_1

or Action = − Reaction.

This is Newton's Third Law of Motion which may be stated as follows:

> To every action there is always an equal and opposite reaction, or the mutual actions of two bodies upon each other are always equal, and directed to contrary parts.

This may be stated in a simpler way as:

> Action and reaction are equal and opposite and act on different bodies.

When a body is attached to the hook of a spring balance, the spring stretches because of the gravitational pull on the body. While the body pulls the spring with a force equal to its weight mg, the spring in turn exerts an equal pull on the body in a direction opposite to that of mg.

Let us consider the example of water drawn from a well by means of a bucket and assume that the rope is passed round a frictionless pulley with a force F applied at its free end.

Consider the forces acting when the bucket is stationary in an elevated position as shown in Fig. 2.19. The weight W of the bucket exerts a pull downwards on the rope. At the same time, the rope exerts an equal and opposite force T on the bucket, which is due to the force exerted by the person holding the rope at the free end.

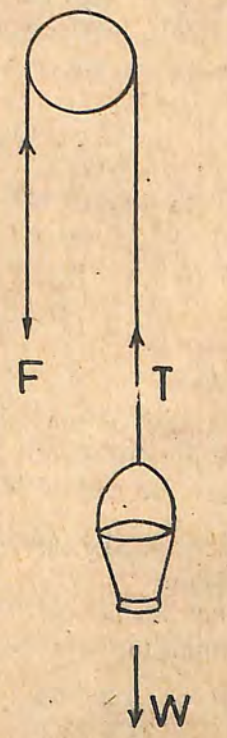


Fig. 2.19 A bucket drawn by passing a rope round a frictionless pulley.

When we apply a force F exceeding the tension needed to balance W, there will be upward motion of the bucket. Under stationary conditions, we find that the pull of the bucket on the rope is balanced by the pull which the rope exerts on the bucket. Also, on the other side of the pulley the downward force applied by the hand on the rope is

counter-balanced by the pull which the rope exerts on the hand. This is an illustration of Newton's Third Law o Motion.

EXAMPLE 2.5

A truck of mass 3000 kg is moving at 10 ms^{-1} and is acted on by two forces, a forward force of 1000 N due to the engine, and a retarding force of 400 N due to friction.

(i) What is the rate at which it is gaining speed?
(ii) How far will it travel in 10 s?

(i) The net force acting on the truck $= 1000 - 400 = 600$ N.
The rate at which it is gaining speed

$$= \frac{F}{m} = \frac{600}{3000} = \frac{1}{5} ms^{-2}.$$

(ii) The distance it will travel in 10 s =

$$10 \times 10 + \frac{1}{2} \times \frac{1}{5} \times 100 = 110 \text{ m}.$$

EXAMPLE 2.6

Two objects of mass 5 kg and 3 kg, respectively are connected by a light cord passing over a horizontal frictionless rod.

(i) What is the acceleration of the system?
(ii) What is the tension in the cord? Assume the value of g to be 9.8 ms^{-2}.

(i) Net force on the system
$= (5 - 3) \times 9.8$ N

Total mass moving $= (5 + 3)$ kg.
Acceleration of the system

$$= \frac{\text{Net force}}{\text{Total mass moving}}$$

$$= \frac{2 \times 9.8}{8}$$

$$= 2.45 ms^{-2}.$$

(ii) Considering the forces on the 5 kg mass, we have

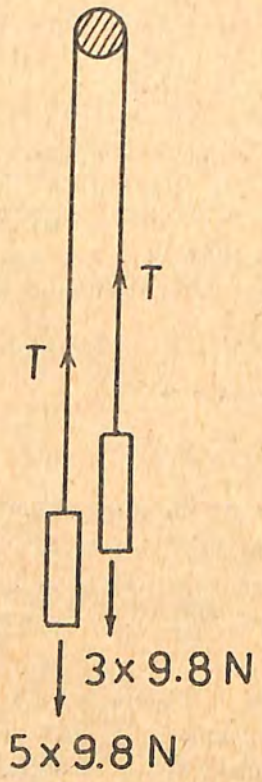


Fig. 2.20 Two weights at the two ends of a light cord, passing over a horizontal frictionless rod.

$$\frac{5 \times 9.8 - T}{5} = 2.45$$

$$\therefore T = 36.75 \text{ N}.$$

The student may work out the value of T considering the forces on the 3 kg mass. In a continuous cord on a smooth surface, the value of T should be the same throughout.

EXAMPLE 2.7

A roller of mass 500 kg is attached by a light horizontal chain to a tractor of mass 1000 kg. The backward force of friction exerted by the ground is 1000 N. If the system has a forward acceleration of 2 ms^{-2}, calculate:

(i) The forward force of the ground on the tractor.
(ii) The tension in the chain.

(i) Considering the entire system, if F is the forward force of the ground on the tractor,

$$\frac{F - 1000}{1500} = 2.$$

$$\therefore F = 4000 \text{ N}.$$

(ii) Considering the tractor alone, if T is the tension in the chain we have,

$$\frac{4000 - T}{1000} = 2$$

$$\text{or } T = 2000 \text{ N}.$$

The student may work out the value of T considering the roller alone.

2.16 Inertial Mass

The same force applied to different bodies produces different accelerations in them. The acceleration produced depends upon the inertia of a body—the greater the inertia, the lesser the acceleration. A measure of the difficulty of accelerating a body is referred to as the *inertial mass* of the body.

If the same force F acting on two bodies produces in them acceleration a_1 and a_2, then

$$\frac{a_2}{a_1} = \frac{m_1}{m_2}$$

where m_1 and m_2 refer to the inertial mass of the bodies.

If one of the bodies is taken as the standard, the above equation can be used to find the inertial mass of the other body in terms of the standard.

Weight

The weight of a body is the force of gravity on the body. If g is the acceleration due to gravity at a place, the weight of a body of mass m is mg at that place. Weight is expressed in units of force. A spring balance gives the weight of a body. Since g varies from place to place, the weight of a body also varies from place to place.

2.17 Collision

We have many examples of collision—carrom-coins colliding with one another, an automobile colliding with another, molecules in a gas colliding with one another, collision of nuclear and atomic particles, etc. When a body in motion collides with another, the velocities of the two bodies change. It is possible to determine their velocities after the collision since the principle of conservation of momentum as well as the principle of conservation of kinetic energy hold good.

During the collision, the transfer of kinetic energy takes place from one body to the other. If there is no loss of kinetic energy during the collision, it is called an *elastic collision*. Otherwise, it is called *inelastic collision*. In actual practice, some kinetic energy is lost in the form of sound and heat during the collision. Hence, the collisions between gross bodies are not perfectly elastic. Certain collisions between atomic and nuclear particles are elastic. Suppose a bullet fired into a sand bag gets embedded in it. Here, the kinetic energy of the bullet is completely used in doing work against the resisting medium. This is an example of a completely inelastic collision. If a body sticks to another body after the collision, it is completely inelastic.

Let a particle of mass m_1 moving with a velocity v_{1i} collide with another particle m_2 at rest (Fig. 2.21). After the collision, let them move with velocities v_{1f} and v_{2f} along directions inclined at θ_1 and θ_2, respectively, with the initial direction of motion of m_1. Let us assume that the collision is elastic.

Taking the scalar components of velocities along the X- and Y-directions and applying the principle of conservation of momentum, we have

Carrom
Billiards Balls.

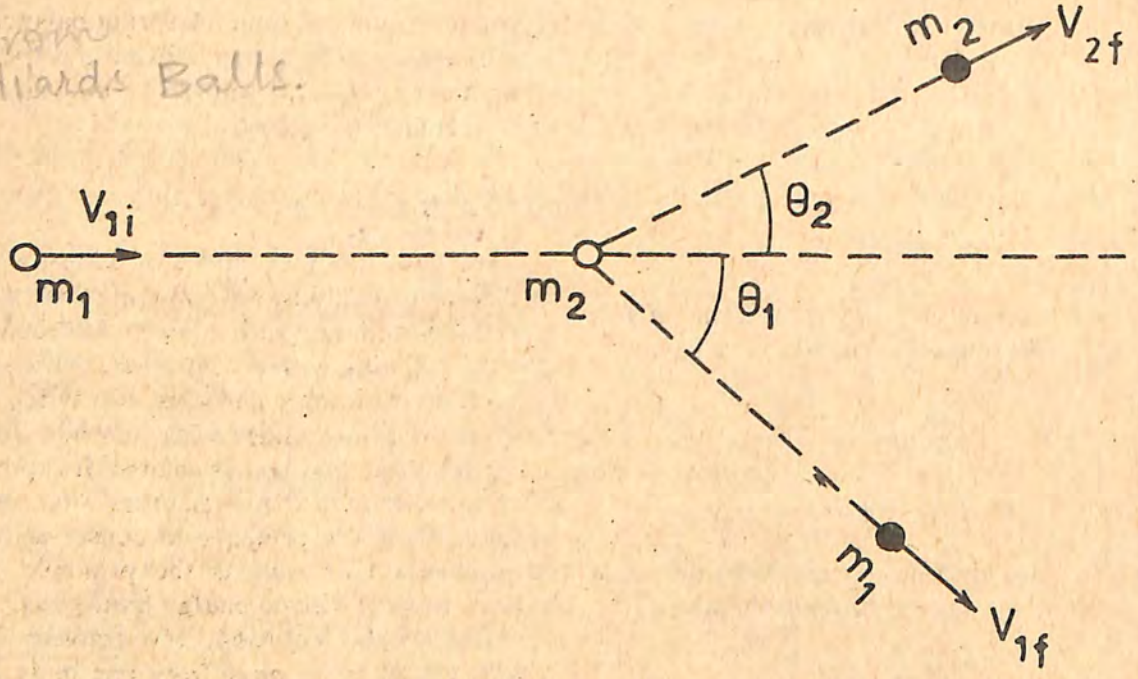


Fig. 2.21 Collision of a particle of mass m_1 and velocity V_{1i} with a particle m_2 at rest. The shaded circles indicate their positions after the collision.

$$m_1 V_{1i} = m_1 V_{1f} \cos\theta_1 + m_2 V_{2f} \cos\theta_2 \quad \ldots (2.16)$$

for the x-component of motion,

and $$0 = m_1 V_{1f} \sin\theta_1 - m_2 V_{2f} \sin\theta_2 \quad \ldots (2.17)$$

for the y-component of motion.

Since the collision is elastic, the kinetic energy is conserved. Therefore,

$$\frac{1}{2} m_1 V_{1i}^2 = \frac{1}{2} m_1 V_{1f}^2 + \frac{1}{2} m_2 V_{2f}^2 \quad \ldots (2.18)$$

Suppose, we know only the values of m_1, m_2 and V_{1i}. We cannot determine the motion after the collision since there are three equations and four unknown quantities, viz., V_{1f}, V_{2f}, θ_1 and θ_2. The value of one more quantity should be known to fix up the motion after the collision. By giving different values for θ_1 or θ_2, the values of the other three unknown quantities can be determined from the above relations.

The Special Case of a Head-on Collision.

Let us consider the special case when $\theta_1 = 0$ and $\theta_2 = 0$. In this case, the particles move along the same direction before and after the collision and it reduces to a case of collision in one dimension. The equations (2.16) and (2.18) will now reduce to

$$m_1 V_{1i} = m_1 V_{1f} + m_2 V_{2f} \quad \ldots (2.19)$$

$$\frac{1}{2} m_1 V_{1i}^2 = \frac{1}{2} m_1 V_{1f}^2 + \frac{1}{2} m_2 V_{2f}^2 \ldots (2.20)$$

On re-arranging and solving for V_{1f} and V_{2f}, we get

$$V_{1f} = \frac{(m_1 - m_2) V_{1i}}{(m_1 + m_2)}$$

and

$$V_{2f} = \frac{2m_1 V_{1i}}{(m_1 + m_2)}.$$

Further, if $m_1 \ll m_2$,

We have

$$V_{1f} = -V_{1i} \quad \text{and} \quad V_{2f} = 0.$$

This means that the lighter particle rebounds from the heavier particle with the same velocity.

EXAMPLE 2.8

A body of mass 1.0 kg makes an elastic collision with another body at rest and continues to move in the original direction with a speed equal to one-fourth of its original speed. Find the mass of the second body.

Incorporating the given data in the equations (2.19) and (2.20), we have:

$$V_{1i} = \frac{1}{4} V_{1i} + m_2 V_{2f}$$

and

$$\frac{1}{2} V_{1i}^2 = \frac{1}{2} \times \frac{1}{16} V_{1i}^2 + \frac{1}{2} m_2 V_{2f}^2$$

From these two equations, we get

$$m_2 = \frac{9}{15} \text{kg}.$$

2.18 Impulse

Let us consider the collision of a body in motion with another body. Let t be the duration of the collision, that is, the time for which the bodies were in contact during which the transfer of momentum has taken place. Let us assume that the direction of the force is constant. Very often we do not know the forces acting during a collision.

Since the rate of change of the momentum of a body is a measure of the force acting on the body, we have

$$\frac{dp}{dt} = F$$

where p and F denote, respectively, the momentum of the body and the force acting on it at any instant of time.

Hence, we have $\int dp = \int F\, dt$. If p_1 is the initial momentum and p_2 the final momentum at the end of an interval of time t, we have

$$p_2 - p_1 = \int_0^t F\, dt.$$

$\int_0^t F\, dt$ is equal to the change in momentum of the body and is a measure of *impulse* of the force. Impulse is denoted by J.

$$\therefore \boxed{J = \int_0^t F\, dt = p_2 - p_1.} \qquad \ldots (2.21)$$

NOTE: The magnitude of the force generally varies during the time of a collision. If $\bar{F}$ is the average value of the force, we have,

$$\boxed{\text{Change in momentum} = \bar{F}.t.}$$

EXAMPLE 2.9

A ball of mass 0.2 kg moving horizontally with a speed of 30 ms^{-1} is struck by a bat. After leaving the bat, the speed of the ball is 40 ms^{-1} in a direction opposite to its original direction of motion. Calculate the impulse of the collision.

The impulse J = Change in momentum of the ball.

$$= (0.2)(-40) - (0.2)(30)$$

$$= -14 \text{ kg ms}^{-1}.$$

Here, we have considered the original direction of motion to be the +ve direction.

2.19 Motion in Two Dimensions—Projectile Motion

Let a particle be projected into the air with some initial velocity. Let us assume

that the effect of the air on the motion of the particle can be neglected. The particle moves in a curved path and the motion will be confined to a vertical plane. The motion of a ball thrown up in the air and that of an object dropped from an aeroplane are examples of projectile motion. The path of the projectile is called its *trajectory*.

Let a body be projected with an initial velocity v_o, making an angle θ with the horizontal ox (Fig. 2.22). Let OX and OY be the coordinate axes with O as the origin in the vertical plane containing the velocity vector.

Since the effect of the air is neglected, the body is subject to only the acceleration due to gravity which acts vertically downwards. The horizontal component of this acceleration is zero. To study the motion of the body, it would be convenient to resolve the velocity of the projectile along the horizontal OX and the vertical OY. The scalar components of v_o along these two directions are, respectively, $v_o \cos \theta$ and $v_o \sin \theta$. Since there is no acceleration along OX, $v_o \cos \theta$, the horizontal component of the velocity remains constant.

$v_o \sin \theta$, the vertical component of the velocity, is directed upwards (Fig. 2.22). This is subject to a downwards acceleration g due to gravity. At any instant of time, the velocity v_y in the vertical direction is given by the relation $v_y = v_o \sin \theta - gt$ where t is the time interval from the start up to that instant of time. As the value of t increases, the vertical component of the velocity decreases in value and becomes zero when t attains the value $\frac{v_o \sin \theta}{g}$. Then the direction

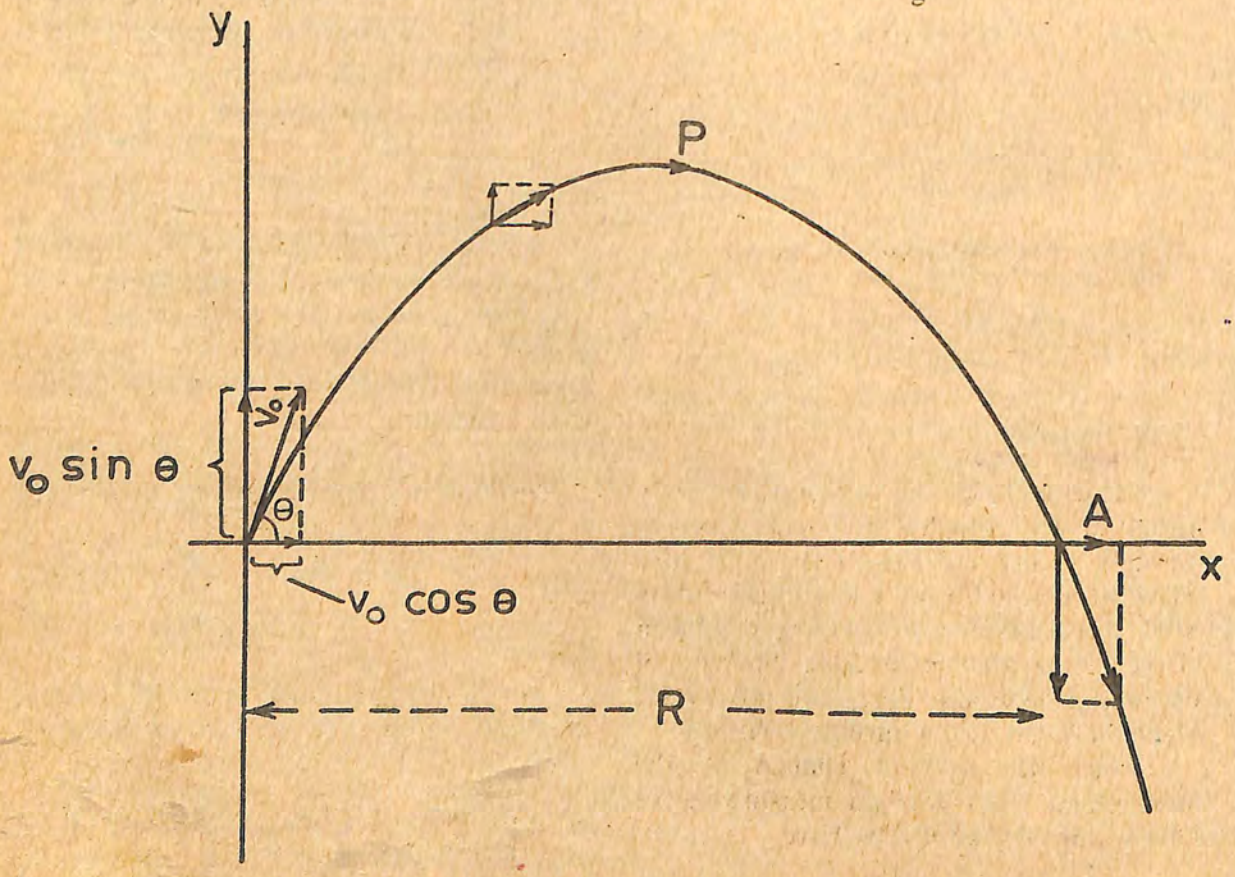


Fig. 2.22 The trajectory of a projectile. Its initial velocity v_o and its scalar components are shown. At P, its velocity is along the horizontal.

of motion of the body will be momentarily horizontal since it has only the horizontal component of velocity at that moment. This corresponds to the highest point P on the trajectory. Afterwards, the vertical component of velocity gradually increases in value in the downward direction until the bullet strikes the ground.

Knowing the velocity of the body along the horizontal and vertical directions at any instant, the resultant velocity can be obtained by compounding these two components. The distances x and y travelled by the body along the horizontal and vertical directions in time t after start can be determined. x and y give the coordinates of the centre of mass of the moving object at any instant of its motion. Thus, the path of the projectile is known.

Equations of the Motion of the Projectile

The velocity along the horizontal OX is:

$$v_x = v_o \cos\theta. \qquad \ldots (2.22)$$

which is constant. The distance moved along the horizontal OX in time t is:

$$x = (v_o \cos\theta)\, t. \qquad \ldots (2.23)$$

The velocity along the vertical Oy at time t from start is

$$v_y = v_o \sin\theta - gt. \qquad \ldots (2.24)$$

The distance moved along the vertical OY in time t is:

$$y = (v_o \sin\theta)\, t - \frac{1}{2} gt^2. \qquad \ldots (2.25)$$

The resultant velocity of the particle is given in magnitude by $\sqrt{v_x^2 + v_y^2}$ and its direction is given by

$$\tan\alpha = \frac{v_y}{v_x}$$

where α is the angle where the direction of the resultant velocity makes with the horizontal OX.

Equation of the Trajectory of the Projectile

Let us consider the equations (2.23) and (2.25) which give us x and y, the coordinates of the projectile at any time t from start.

On eliminating t between the two equations, we arrive at an equation of the form

$$y = px - qx^2 \qquad \ldots (2.26)$$

where p and q are constants.

This is the equation of a parabola.

Therefore, the trajectory of a projectile is a parabola.

2.20 The Range of a Projectile

In Fig. 2.22, O is the point of projection and A is the point where the trajectory cuts the horizontal through O. The distance OA is called the *horizontal range* or simply the range of the projectile and is denoted by R.

At the point A, the y coordinate is zero. So, we have

$$O = (v_o \sin\theta)\, t - \frac{1}{2} gt^2$$

from which

$$t = \frac{2v_o \sin\theta}{g}.$$

The range R is the horizontal distance travelled by the projectile in this time interval t.

$$\therefore R = (v_o \cos\theta)\left(\frac{2v_o \sin\theta}{g}\right)$$

$$R = \frac{v_o^2 \sin 2\theta}{g}. \qquad \ldots (2.27)$$

From the expression for R, it follows that, for a given velocity of projection, the value of R is maximum when sin 2θ is maximum, i.e., when $2\theta = 90°$ or $\theta = 45°$.

EXAMPLE 2.10

A shell is fired horizontally from a gun situated at a height of 44.1 m above a horizontal plane with a muzzle velocity of 300 ms^{-1} to hit a target on the horizontal plane.

(i) How long does the shell take to hit the target?
(ii) Where is the target situated?
(iii) With what vertical velocity does the shell strike the target?

Given $g = 9.8\ ms^{-2}$

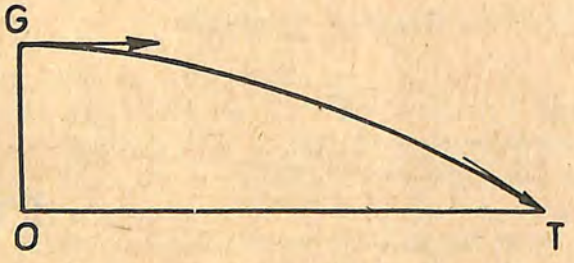


Fig. 2.23 A horizontally fired shell OG = 44.1 m.

(i) Initially, the vertical component of velocity is zero. If t is the time taken by the shell to hit the target, we have from the equation (2.10)

$$44.1 = \frac{1}{2} \times 9.8 \times t^2$$

From which $t = 3$ seconds.

(ii) T is the target (Fig. 2.23) and OT is the range of the projectile.

OT = (Horizontal component of velocity) (time)

$= 300 \times 3$

$= 900$ m.

(iii) If the vertical component of the velocity of the shell is v_y when it hits the target, from the equation 2.11 we have:

$$v_y^2 = 2 \times 9.8 \times 44.1$$

$$\therefore v_y = 29.4\ ms^{-1}.$$

2.21 Work, Energy and Power

We have learnt in the earlier classes the equivalence between work and energy. We have known that there are two forms of mechanical energy—kinetic and potential. Kinetic energy is the energy of motion and potential energy is the energy due to the state of strain of a body or due to its relative position with respect to another body with which it interacts (Example: gravitational potential energy). Whenever work is done by a force acting on a body, the energy of the body changes.

2.22 Work

The amount of work done by a force is given by the product of the force and the distance through which the point of application of the force moves in the direction of the force.

If θ is the angle between the directions of the force F and displacement S the magnitude of the component of F parallel to S is $|F| \cos\theta$, or simply $F \cos\theta$ (Fig. 2.24). The work done is given by:

$$W = (F \cos\theta)(S). \qquad \ldots (2.28)$$

When $\theta = 0$, the force is in the same direction as the displacement. In this case, $\cos\theta = 1$ and we have $W = F.S$.

When $\theta = 90°$, $\cos\theta = 0$ and no work is done by the force even though there is a shift in the point of application of the force. For example, in the case of the moon going round the earth, no work is done by the centripetal

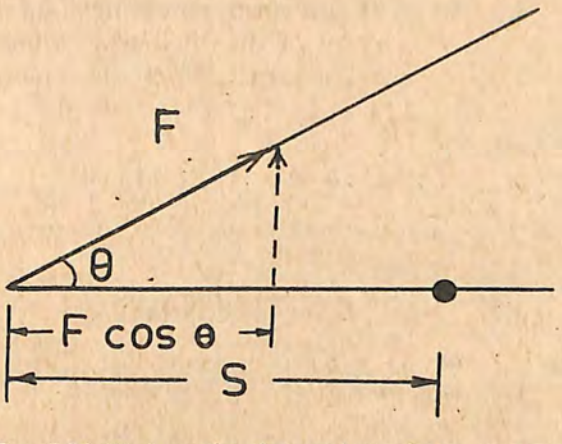


Fig. 2.24 Work as the dot product of two vectors: **F** and **S**. W = **F.S** = (F cos θ) S.

force since the motion of the moon is throughout at right angles to the direction of the force.

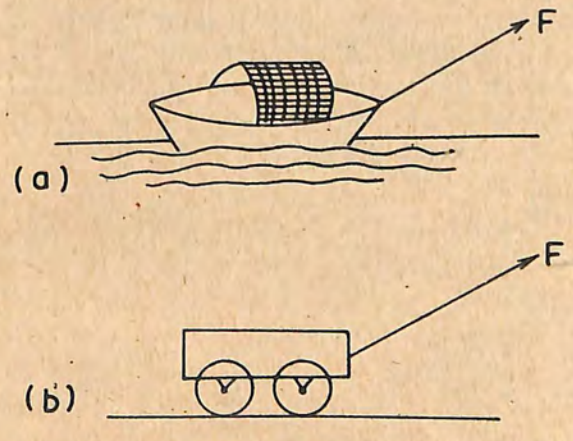


Fig. 2.25 (a) A boat in a canal pulled by a rope; (b) A cart on a road pulled by a rope.

Fig. 2.25a shows a boat in a canal pulled by a rope and Fig. 2.25b shows a cart on a road also pulled by a rope. In either case, the force is inclined to the horizontal direction, while the point of application of the force is shifted along the horizontal. In both the cases, the work done is given by the equation (2.28).

Work is a scalar and we may consider work as the dot product of the two vectors F and S.

$$W = F.S \qquad \ldots (2.29)$$

A unit work is done when the point of application of a force of 1 N is shifted through 1 m in the direction of the force. It is called *newton-metre* or *joule*.

In the cgs system, a unit work is done when the point of application of a force of 1 dyne is shifted through 1 cm in the direction of the force. It is called *erg*.

$$1 \text{ joule} = 10^7 \text{ ergs}.$$

2.23 Work Done by a Constant Force

Let a constant force F act on a body and displace it through s in the direction of the force. The work done W is given by

$$W = F s.$$

The total work W for a shift s is given by

$$\int dW = \int F\, ds.$$

We have $F = ma = m \dfrac{dv}{dt}$

and $\dfrac{dv}{dt}$ can be written as $\dfrac{dv}{ds} \cdot \dfrac{ds}{dt}$ so that

$$W = \int m \left(\frac{dv}{ds} \cdot \frac{ds}{dt} \right) ds.$$

On replacing $\dfrac{ds}{dt}$ by v, the above integral reduces to

$$W = \int mv\, dv.$$

If v_i is the initial velocity of the body before the application of the force and v_f its final velocity,

$$W = \int_{v_i}^{v_f} mv\, dv = \frac{1}{2} m v_f^2 - \frac{1}{2} m v_i^2.$$

Since W = Fs, the above equation can be written as:

$$Fs = \frac{1}{2} m v_f^2 - \frac{1}{2} m v_i^2. \qquad \ldots (2.30)$$

If the body is initially at rest, $v_i = 0$ and the work done by the force or the energy transferred will be $\frac{1}{2}$ m v_f^2. This expression is a function of velocity and is a measure of the kinetic energy of the body.

The right-hand side expression of the equation (2.30) gives the change in the kinetic energy of the body. Therefore, the work done is equal to the change in the kinetic energy of the body.

Let us now apply this to the case of a freely falling body.

Let a body of mass *m* be released from a height *h* above the surface of the earth. For values of *h* which are very small compared to the radius of the earth, the variation in the acceleration due to gravity can be neglected. Therefore, the force of gravity mg acting on the body can be considered as constant. The work done by the force of gravity in shifting the body through a distance *h* is equal to mgh. Since the distance of the body from the earth has decreased by *h*, there is a loss in the potential energy of the body relative to the earth. This loss in P.E. is equal to mgh, the work done by the force of gravity.

If v is the velocity acquired by the body on dropping through *h*, we have $v^2 = 2gh$ from the equation (2.11). Hence, mgh = $\frac{1}{2}$ mv^2, which is the kinetic energy of the body. Therefore, the loss in the potential energy is equal to the gain in the kinetic energy of the body.

EXAMPLE 2.11

An engine pulls a train 1 km over a level track. Calculate the work done, given that the frictional resistance is 5×10^5 N.

Work done = (Force) (Distance moved in the direction of the force.)

The train exerts a force equal and opposite to the frictional resistance.

$\therefore$ Work done = $5 \times 10^5 \times 10^3$ joules.

= 5×10^8 joules.

EXAMPLE 2.12

A neutron having mass of 1.67×10^{-27} kg is moving at a speed of 2×10^4 ms^{-1}. What is its kinetic energy?

The kinetic energy of the neutron

$$= \frac{1}{2} \times 1.67 \times 10^{-27} \times (2 \times 10^4)^2$$

$$= 3.34 \times 10^{-19} \text{ joules.}$$

EXAMPLE 2.13

A boy of mass 50 kg sits in a swing suspended by a rope 5 m long. A person pulls the swing to a side so that the rope makes an angle of 30° with the vertical. What is the gain in the gravitational potential energy of the boy? The acceleration due to gravity at the place is 9.8 ms^{-2}.

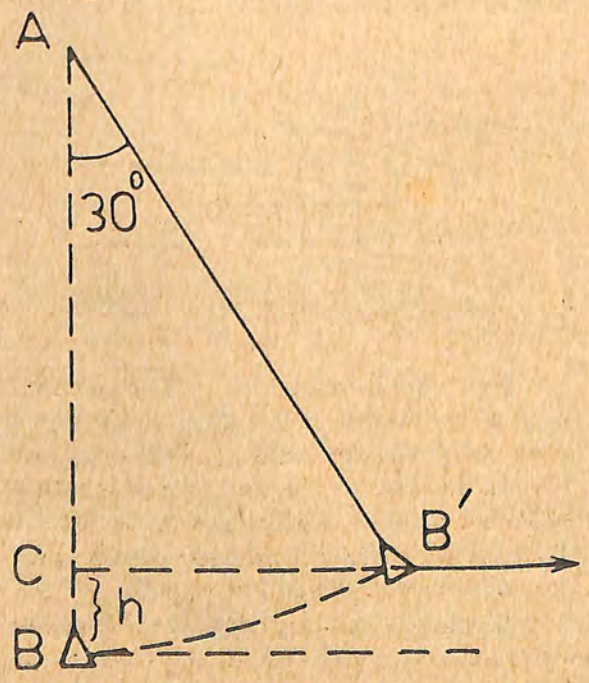


Fig. 2.26 A swing drawn to a side (the potential energy of the swing in relation to the earth increases): **AB = AB′** = 5 m.

The height h through which the boy is raised

$= AB - AC$

$= 5 - 5\cos 30^\circ = 5(1 - \cos 30^\circ)$

$= 5\left(1 - \sqrt{\frac{3}{2}}\right) = 0.670$ m.

The gain in P.E. of the boy $= mgh$,

$= 50 \times 9.8 \times 0.67$

$= 328.3$ joules.

EXAMPLE 2.14

A bullet of mass 0.03 kg moving with a speed of 400 ms^{-1} penetrates 12 cm into a fixed block of wood. Calculate the average force exerted by the wood on the bullet.

The kinetic energy of the bullet just before striking the block of wood is:

$\frac{1}{2} \times 0.03 \times (400)^2$ joules.

The entire kinetic energy is used in doing work against the average resisting force F of the wood.

$\therefore F \times 0.12 = \frac{1}{2} \times 0.03 \times (400)^2$.

$\therefore F = 2 \times 10^4$ N.

EXAMPLE 2.15

An automobile moving at a speed of 72 kmh^{-1} reaches the foot of a smooth incline when the engine is switched off. How much distance does the automobile go up the incline before coming to rest. The inclination of the plane to the horizontal is 30° and the acceleration due to gravity at the place is 9.8 ms^{-2}. The speed of the automobile $= 72\ km\ h^{-1}$

$= \frac{72 \times 10^3}{60 \times 60} = 20\ ms^{-1}$.

The kinetic energy of the automobile just before climbing the incline is:

$\frac{1}{2} m (20)^2$ joules.

where mkg is the mass of the automobile. This energy is used completely to do work against the component of the gravitational force on the automobile down the plane.

The component of the gravitational force down the plane is:

$mg \sin \theta$

$= m \times 9.8 \times \sin 30^\circ$.

If a car goes a distance S along the plane before coming to rest, we have

$(m \times 9.8 \times \sin 30^\circ) S = \frac{1}{2} m (20)^2$

$\therefore S = 40.8$ m.

2.24 Work Done Against a Variable Force

In actual situations, the magnitude of the force acting on a body may not always remain constant. For example, the stress developed in an elastic string increases with extension within limits of elasticity. To elongate a string, work has to be done against the stress developed in the string.

A graphical method can be adopted for estimating the work done when the force is variable. For simplicity, let us assume that the displacement is parallel to force. Fig. 2.27 gives a force displacement curve when the force is variable.

Consider a very small displacement ds_1, as shown in Fig. 2.27, after OP where OP = S. Let F_1 be the magnitude of the force acting over this small displacement ds_1. F_1 is almost constant over this displacement ds_1; W, the work done for the displacement ds_1, is given by $F_1 . ds_1$. The area of the thin-shaped rectangle represents the

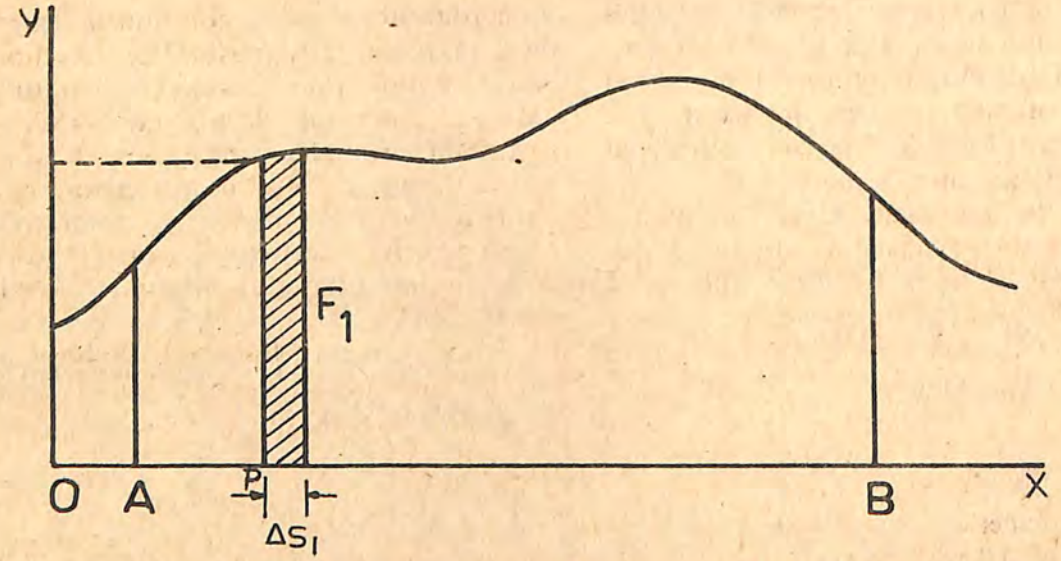


Fig. 2.27 A force-displacement curve. OX—displacement, OY—force.

work $F_1 . ds_1$. Suppose, we are required to estimate the work done over a displacement AB. The entire displacement AB can be divided into a large number of small displacements. The total work done for the displacement AB will be given by the sum of the areas of all the rectangles erected on all the small displacements. In the limit when $S \to 0$ the number of rectangles tends to infinity and the sum of the areas of all the rectangles will approach the area under the force displacement curve (i.e., the area bounded by the curve, the displacement axis and the ordinates at A and B). Therefore, the area of figure gives the work done. In calculus notation we may write,

$$W = \int_{S=OA}^{S=OB} F . ds. \qquad \ldots (2.31)$$

2.25 Energy

We have various forms of energy, like mechanical, heat, light, sound, chemical and mass energy.

The primitive man produced fire by striking one flint against another. It is common experience that on a cold winter day we often rub one hand against the other. Thereby, we generate some heat. When a body is moved on a rough surface, work is done against the frictional resistance and the body gets heated. In all these cases, mechanical energy is converted into heat energy.

Let a ball be dropped from a certain height above the surface of the earth. If the collision is inelastic, the ball will not rise to the same height from which it is dropped after hitting the ground. In such a case, the magnitude of the velocity after the collison will be less than that before the collision. Part of the kinetic energy of the body is dissipated during the impact predominantly in the form of heat energy and to some extent in the form of sound energy also. When coal is burnt, the chemical energy of coal is released as heat energy. This heat energy can be used to heat water and produce steam which in turn can be made to run a steam engine. In this case, the chemical energy is first transformed into heat energy and then into mechanical energy.

In the generation of hydro-electricity, the mechanical energy of water stored at a height is transformed into electrical energy.

Hydro-electric stations are built at places where a river has a steep gradient or near a water-fall. A dam is constructed just ahead of the water-fall to store the water. The flow of water from the reservoir is regulated through large pipes leading down to the foot of the water-fall where the hydro-electric station is constructed. The water flowing out of the pipes with terrific speed hits the blades of a water turbine and sets it in rapid rotation (Fig. 2.28). The turbine is coupled with the rotor of a dynamo and electricity is generated.

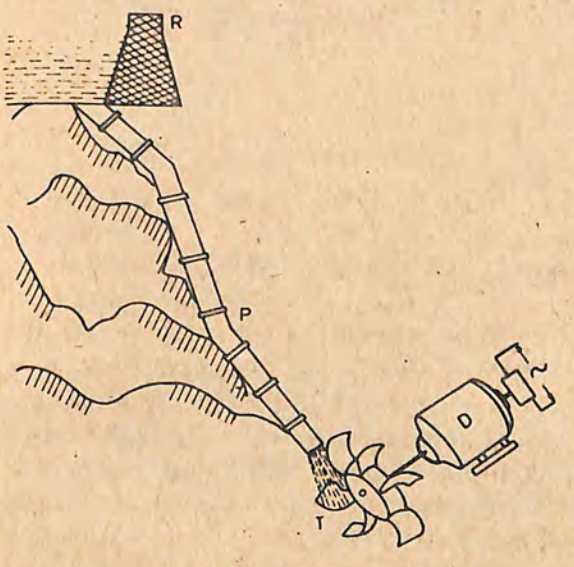


Fig. 2.28 The principle underlying the generation of hydro-electricity: R—reservoir; P—penstock pipes; D—dynamo; T—water turbine.

We have now many hydro-electric stations in India such as Bhakra-Nangal project, Baira-Siul Hydro-electric project, Machkund and Sileru projects, Kundah project, Srisailam project and Mettur project. Compared with the cost of generation of electricity using coal or mineral oils, the cost of production of hydro-electric power is cheap. The coal and mineral oil resources of any country are limited whereas it is not so with water power. This is an important consideration in favour of hydro-electricity.

By means of high tension cables, electricity is transmitted to the consuming centres where it is put to various uses. The household uses of electricity are mainly for lighting, heating and running electric motors as in the case of fans. These three uses illustrate the transformation of electrical energy into light, heat and mechanical energy, respectively. In factories, electricity is mainly used for driving machinery by means of electric motors.

Einstein has established (1905) the equivalence between mass and energy and stated that they are related by the equation

$$E = mc^2 \qquad \ldots (2.32)$$

where E is energy, m is mass and c is the speed of light. Since c is 3×10^8 ms^{-1}, the energy equivalent to even a small mass of 1 g is $(10^{-3})(3 \times 10^8)^2$ or 9×10^{13} joules, which is a very large value. In this universe, most of the energy is in the form of mass energy. But we do not know how to convert mass completely into energy, though during nuclear reactions, including fusion and fission reactions a very small proportion of mass is converted into energy. The energy which the sun continuously emits is derived as a result of the conversion of a small quantity of mass into energy during the fusion of Hydrogen nuclei taking place continuously in the sun.

2.26 Conservation of Energy

We have seen that when energy in one form disappears, it reappears in another form. Energy is not lost or destroyed. The total energy in a system is conserved. We have the principle of conservation of energy which is stated as follows:

> Energy cannot be created or destroyed. If may be transformed from the kind to another. The total energy in a system is constant.

This is a fundamental law or basic

principle in physics like the principle of conservation of momentum.

2.27 Power

In pulling a road roller over a certain distance some amount of work is done. Whether it is pulled in 2 hours or 10 hours, the total work done is the same. But the rate of doing work differs. *The time rate at which work is done is called power.*

The unit of power in the mks system is the work done at the rate of 1 joule per second. This is called *watt* (symbol W) named after James Watt, the inventor of the steam-engine.

In the cgs system the unit of power is 1 erg per second. There is another unit of power used in practice which is called the *horse-power*. One horse-power is equal to 746 watts.

If the rate at which work is done is constant, then Work done in time t = (power) × t.

The rating of electric lamps is generally given in watts. A 40 watt lamp consumes electrical energy at the rate of 40 joules per second. If this lamp is lighted for an hour, the energy consumed is 40 × 60 × 60 joules. This is also expressed in (units of power) × (time). In this case, the energy consumed is stated as 40 watt-hours. It should be noted that the watt-hour is a unit for energy but not for power. Electrical energy consumed is measured in terms of the kilo-watt-hour which is 1000 watt-hours.

$$1 \text{ kilo-watt hour} = 1000 \times 60 \times 60$$
$$= 36 \times 10^5 \text{ joules.}$$

EXAMPLE 2.16

A house is fitted with 5 electric lamps, each of 40 watt capacity, and other 5 lamps of 60-watt capacity each. If all the lamps are lighted (at their maximum capacity) for 8 hours in a day, what is the consumption of electrical energy in a day?

Electrical energy consumed in a day is:

$$(5 \times 40 + 5 \times 60) \times 8 \text{ watt-hours}$$
$$= 4000 \text{ watt-hours}$$
$$= 4 \text{ kilo-watt} - \text{hours.}$$

EXAMPLE 2.17

A motorboat is moving at a steady speed of 20 ms^{-1}. If the water resistance of the motion of the boat is 6000 N, calculate the power of the engine.

Power of the engine is:

The rate at which work is done.

$$= 6000 \times 20 \text{ joules per second}$$
$$= 12 \times 10^4 \text{ watts.}$$

EXAMPLE 2.18

A pump-set is used to lift water to a reservoir of 6000 litres capacity over an average height of 20 m. If it takes 1 *h* to fill the reservoir completely, calculate the work done and the power supplied to the set if its efficiency is 60 per cent. The acceleration due to gravity is 9.8 ms^{-2}.

Mass of 1 litre of water = 1 kg.

∴ The mass of water lifted = 6000 kg.

Total work done = $6000 \times 9.8 \times 20$.

$$= 11.76 \times 10^5 \text{ joules.}$$

Since the efficiency is only 60 per cent, the power supplied to the set

$$= (\text{work done}) \times \frac{100}{60} \times \frac{1}{(\text{time in seconds})}$$
$$= \frac{11.76 \times 10^5 \times 100}{60 \times 60 \times 60} = 544.4 \text{ W}$$

2.28 Friction

We have discussed under the laws of motion that in nature we do not come across a perfectly smooth surface. As a near approximation we have considered the 'air track' wherein a cushion of air is provided for the cart to move, thus eliminating, to a great extent, the friction between the surfaces.

Consider a block on a horizontal surface. Let it be given a push and released. The velocity goes on decreasing and it finally comes to rest. The retardation of the block is due to the force of friction between the surfaces in contact. The magnitude of this force of friction depends upon the nature of the surface in contact. Friction is in a way helpful to us. We are able to walk because of the friction between the feet and the ground. In the absence of friction as on a perfectly smooth surface, we cannot walk. We slip and fall down. Everyone must have exeperienced at one time or other slipping on a smooth floor, especially when it is oily.

2.29 Origin of Friction

Even a highly polished surface when seen under a powerful microscope reveals dents and irregularities of the surface (Fig. 2.29).

Fig. 2.29 The surfaces of two bodies in contact. The figure shows how the magnified image of the section looks.

When a pull is applied on one body so as to cause its surface to slide over the other surface, the relative motion, on account of the large number of surface irregularities, is resisted and an opposing force develops. This is the force of friction. As the pull increases, the resisting force also increases but at a certain stage the body slides.

2.30 Nature of Friction

Whenever the surface of one body slides over the surface of another body, each body exerts a frictional force on the other parallel to the surfaces. The frictional force on each body is in a direction opposite to its direction of motion relative to the other body.

In the example of the block over a horizontal rough surface, if the block slides in the direction AB (Fig 2.30a), the force of friction acts in a direction opposite to AB.

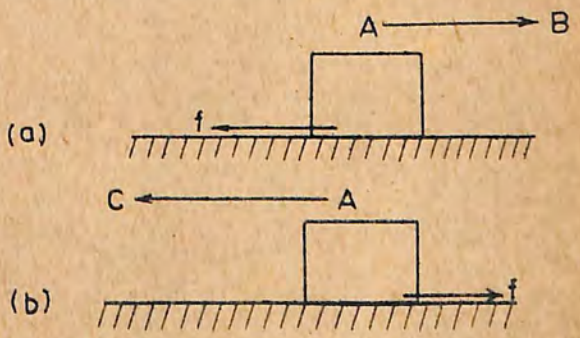


Fig. 2.30 A block over the horizontal rough surface. The force of friction acts in a direction opposite to the direction in which the body slides

If the direction of motion is reversed and the block slides in the direction of AC (Fig. 2.30b), the force of friction f gets reversed and acts opposite to AC. Therefore, friction always opposes motion.

Suppose, a block is resting on a horizontal table. Let a small force f_1 be applied to it in the direction AB by means of a spring

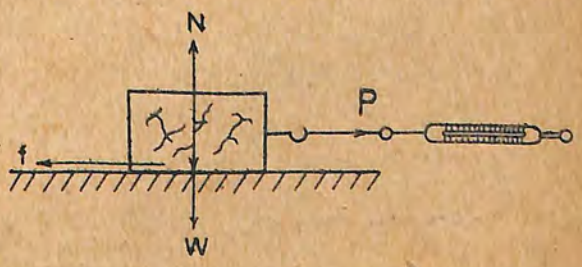


Fig. 2.31 A block pulled by a spring balance. As the pull applied to the block increases, the magnitude of the force of friction increases and maintains the body at rest until the force of friction reaches its limiting value known as the maximum force of static friction

balance (Fig. 2.31). If the block does not move, it means a frictional force is developed which exactly opposes f_1 and cancels it. This is called *static friction*. Thus, even when there is no relative motion, frictional force acts.

Let the pull on the block be gradually increased by pulling the spring balance. The magnitude of the force of friction also increases and maintains the body at rest until at a certain stage the force of static friction reaches its limiting value (maximum value). When the pull P applied to the block just exceeds the maximum force of static friction f_{ms} (developed in the opposite direction), the block starts moving. As the block slides, the magnitude of the frictional force decreases and the new value of the force of friction is called the *force of kinetic friction* (f_k), sometimes also called the *force of sliding friction*. f_k is less than f_{ms}. Therefore, as the block moves, the net force acting is $f_{ms} - f_k$. This resultant force will cause some acceleration. If now the value of P is decreased so that it equals f_k, the block will move with uniform speed.

It may be noted that the work done in overcoming friction which is equal to Σf_k. Δs is converted into heat and does not produce kinetic energy in the block.

The weight W of the block acts vertically downwards. Since there is no motion in the vertical direction, the normal reaction exerted on the block by the table just balances the weight of the body. So, the normal force N which acts vertically upwards is equal in magnitude to the weight of the body.

2.31 Laws of Friction

From the various experiments relating to 'dry friction' (that is, when the surfaces are dry and unlubricated), the following two empirical laws are found to be obeyed by the maximum force of static friction:

(i) The maximum force of static friction is almost independent of the area of contact, and

(ii) the maximum force of static friction is proportional to the normal force.

The ratio of the maximum force of static friction to the normal force is called the *coefficient of static friction* for the pair of surfaces concerned. It is generally denoted by μ_s. If f_{ms} is the maximum force of static friction and N is the normal force,

$$\frac{f_{ms}}{N} = \mu_s$$

$$\text{or } f_{ms} = \mu_s N. \qquad \ldots (2.33)$$

For values of P less than f_{ms}, the force of static friction developed is less than the limiting value, that is, less than μ_s N. So, we can write

$$f_s \leq \mu_s N.$$

Kinetic Friction

For kinetic friction between dry unlubricated surfaces, there are two empirical laws similar to those of static friction.

The ratio of the force of kinetic friction to the normal force is called the *coefficient of kinetic friction*, or *co-efficient of sliding friction*. It is denoted by μ_k and

$$f_k = \mu_k N. \qquad \ldots (2.34)$$

TABLE 2.1

Coefficients of Friction (approximate)*

Surfaces	Condition	Coefficient of friction	
		Kinetic	Static
Steel on steel	dry	0.18	0.25
Steel on wood	dry	0.20	0.40
Wood on wood	dry	0.25	0.50
Leather on wood	dry	0.40	0.55
Stone on concrete	dry	0.45	0.75
Steel on steel	greased	0.05	0.10
Car tyre on concrete	moderate speed	0.40	0.60

* These are only typical values. They vary greatly depending upon the condition of surfaces, grain of wood, contamination, moisture, etc.

Since $f_{ms} > f_k$, we have $\mu_s > \mu_k$; μ_s and μ_k are constants depending upon the nature of the surfaces.

EXAMPLE 2.19

A box of mass 4 kg rests upon an inclined plane. The inclination of the plane to the horizontal direction is gradually increased. It is found that when the slope of the plane is 1 in 3, the box starts sliding down the plane. Given, $g = 9.8\ ms^{-2}$.

(i) Find the coefficient of friction between the box and the plane.

(ii) What force applied to the box parallel to the plane will just make it move up the plane?
It is given that the slope of the plane is 1 in 3.

It means that

$$\frac{\text{Height BC of the plane}}{\text{Length AC of the plane}} = \frac{1}{3}$$

That is, $\sin\theta = \frac{1}{3}$

The weight W may be resolved into W $\sin\theta$ along the plane and W $\cos\theta$ perpendicular to the plane.

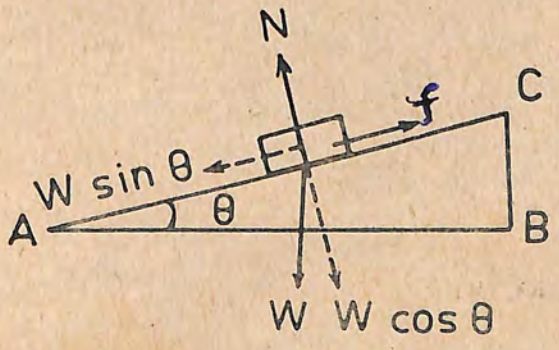


Fig. 2.32 A box on an inclined plane. When the sliding commences, $\tan\theta = \mu$ (F is the force of friction and N is the normal force).

(i) Let N be the normal force and f the force of friction. For the equilibrium of the forces, when the box just begins to slide.

$$f = W \sin\theta. \qquad \ldots (2.35)$$

$$N = W \cos\theta. \qquad \ldots (2.36)$$

$$\therefore \frac{f}{N} = \tan\theta = \frac{1}{\sqrt{3^2 - 1^2}}$$

$$\mu = \frac{1}{2\sqrt{2}}$$

(ii) Let F be the minimum force needed to move the box up the plane. Then $F = W \sin\theta + f$,

since the force of friction f now acts down the plane.

Substituting for f from the equation (2.35),

$$F = W \sin\theta + W \sin\theta = 2\,W \sin\theta$$

$$= 2 \times 4g \times \frac{1}{3} = \frac{8}{3} \times 9.8$$

$$= 26.13\ N$$

We have seen in the above example that as the angle of inclination of the plane to the horizontal is gradually increased, for a particular value of θ sliding commences. This value of the angle of inclination is called the *angle of static friction*. Its value naturally depends upon the nature of the surfaces and it may be noted that $\tan\theta = \mu$.

2.32 Rolling Friction

When one body rolls over another, the frictional force developed is called the *force of rolling friction* and the corresponding coefficient is called the *coefficient of rolling friction* (μ_r). μ_r is found to be much less than μ_s for the same two surfaces. Since the rolling friction is much smaller than the sliding friction, the wheel has come into so much use in everyday life. It is much easier and less

laborious to pull a loaded cart supported on wheels rather than to draw the cart allowing it to slide over the ground.

2.33 Lubrication

It is mentioned at the beginning of the sub-unit 'Friction' that we are able to walk and move about because of friction between the feet and the ground. Two pieces of wood are held together by means of nails due to friction between the nails and wood. When brakes are applied to a bicycle in motion, the brake blocks touch the wheel and the force of friction between the wheels and brake blocks reduces the speed of the bicycle.

While friction is helpful in some cases, it acts as a hindrance in other cases. In a machine, if there is friction between the moving parts, a lot of energy that is applied to the machine is transformed into heat energy while overcoming the frictional resistance. The heat that is produced may cause serious damage to the machine, apart from causing wear and tear of the moving parts.

Friction can be reduced considerably by using suitable surfaces and by lubrication. It is already mentioned that the rolling friction is much smaller than the sliding friction. This is taken advantage of in the moving parts of machinery by making the surfaces hard and by keeping hard rollers or balls between them. If the metal surfaces of ball bearings are not hard, then the friction will be high. That is why hard steel balls are used in ball bearings (Fig. 2.33).

(a) (b)

Fig. 2.33 (a) Ball bearing; (b) A section showing the relative motion of the parts in a ball bearing.

In a big machine, a continuous supply of oil or some other suitable lubricant is provided for the smooth functioning of the machine and to prevent the moving parts from getting unduly heated. Surfaces slide on grease or oil film more easily than on dry surfaces. This helps smoother running and reduction in the wearing out of the moving parts.

The use of air as a lubricant has great potentialities. Purified air is compressed and employed as a lubricant. It forms an elastic cushion between the moving parts and eliminates the problem of friction and heating. The outflow of air from the machine also prevents the entry of dust into the machine. The principle of floating a vehicle on a cushion of air from a compressor to prevent ground friction is used for efficient locomotion in Hoover craft.

EXERCISES

2.1 Distinguish between vector and scalar quantities. Give examples.
2.2 Can three vectors not in one plane give a zero resultant? Can four vectors do?
2.3 If $\mathbf{A} \times \mathbf{B} = \mathbf{C} \times \mathbf{B}$, show that $\mathbf{C}$ need not be equal to $\mathbf{A}$. When will $\mathbf{A}$ be equal to $\mathbf{C}$?
2.4 Are the commutative and associative laws applicable to vector subtraction?
2.5 Is it possible for a body to be accelerated, if its speed is constant?
2.6 Under what conditions will the pull of a lift (mass being fixed) on the cable be a maximum and a minimum?

2.7 A goods train has 25 compartments. Will the tension in the coupling between the fourth and the fifth compartment be the same as that in the coupling between the twenty-first and the twenty-second compartment?

2.8 How do you compare the inertial mass of two bodies?

2.9 The magnitude of the velocity of projection being fixed how does the range of a projectile depend upon its angle of projection? Is there any optimum value for the angle of projection so that the range may be a maximum?

2.10 In a thermal station, coal is used for the generation of electricity. Mention how energy changes from one form to another before it is transformed into electrical energy.

2.11 Give a few instances where friction is helpful to us.

2.12 What are the factors on which the coefficient of friction between two surfaces depends? How does lubrication help to minimize friction?

2.13 An aeroplane takes off at an angle of 30° to the horizontal. If the component of its velocity along the horizontal is 250 kmh^{-1}, what is its actual velocity? Find also the vertical component of its velocity.

(288.7 kmh^{-1}; 144.4 kmh^{-1})

2.14 A particle has a displacement of 12 m towards the east and 5 m towards the north and then 6 m vertically upward. Find the magnitude of the sum of these displacements.

(14.32 m)

2.15 A ship is sailing westward at 8 m/s. While trying to fix a bolt at the top of the mast, a sailor drops the bolt. If the mast of the ship is 19.6 m high, where will the bolt hit the deck?

2.16 A body is dropped from rest at a height of 150 m and, simultaneously, another body is dropped from rest from a point 100 m above the ground. What is their difference in height after they have fallen (a) 2 s (b) 3 s. How does the difference in height vary with time?

(50 m)

2.17 A balloon is ascending at the rate of 14 m/s at a height of 98 m above the ground when a packet is dropped from the balloon. After how much time and with what velocity does it reach the ground?

($6^1/_8$ s; 46.02 ms^{-1})

2.18 A parachutist bails out from an aeroplane and after dropping through a distance of 40 m he opens the parachute and decelerates at 2 ms^{-2}. If he reaches the ground with a speed of 2 ms^{-1}, how long is he in the air? At what height did he bail out from the plane?

($15\frac{8}{7}$ s; 235 m)

2.19 Fuel is consumed at the rate of 100 kg per second in a rocket. The exhaust gases are ejected at a speed of 4.5×10^4 ms^{-1}. What is the thrust experienced by the rocket?

(4.5×10^6 N)

2.20 An elevator weight 4000 kg. When the upward tension in the supporting cable is 48000 N, what is the upward acceleration? Starting from rest, how far does it rise in 3 seconds?

(2.2 ms^{-2}; 9.9 m)

2.21 A boy whose mass is 50 kg stands on a spring weighing machine inside a lift. The lift starts to ascend with an acceleration of 2.45 ms^{-2}. What is the reading in the machine? What will be the reading when in its ascent the lift moves (a) with uniform velocity, (b) with the retardation of 4.9 ms^{-2}? (c) falls freely.

(62.5 kg; 50 kg; 25 kg)

2.22 A neutron having mass of 1.67×10^{-27} kg and moving at 10^8 ms^{-1} collides with a deuteron at rest and sticks to it. Given that the mass of the deuteron is 3.34×10^{-27} kg, calculate the speed of the combination. (The composite particle is called triton.)

(1/3 10^8 ms^{-1})

2.23 A ball is dropped from rest at a height of 12 m. If it loses 25 per cent of its kinetic energy on striking the ground, what is the height to which it bounces? How do you account for the loss in K.E.?

(9 m)

2.24 While launching a rocket of mass 2×10^4 kg, a force of 5×10^5 N is applied for 20 s. What is the velocity attained by the rocket at the end of 20 seconds?

(500 ms^{-1})

2.25 From the top of a building 19.6 m high, a ball is projected horizontally. After how long does it strike the ground? If the line joining the point of projection to the point where it hits the ground makes an angle of 45° with the horizontal, what is the initial velocity of the ball?

(2 s; 9.8 ms^{-1})

2.26 A bomb is dropped from an aeroplane when it is directly above a target at a height of 1500 m. The aeroplane is going horizontally with a speed of 500 km h^{-1}. By how much distance will the bomb miss the target?

(2431 m)

2.27 A bullet of mass 0.01 kg is fired horizontally into a 4 kg wooden block at rest on a horizontal surface. The coefficient of the kinetic friction between the block and the surface is 0.25. The bullet gets embedded in the block and the combination moves 20 m before coming to rest. With what speed did the bullet strike the block?

(198.3 ms^{-1})

2.28 A truck tows a trailer of mass 1200 kg at a steady speed of 10 ms^{-1} on a level road. The tension in the coupling is 1000 N. What is the power expended on the trailer?

Find the tension in the coupling when the truck ascends a road having an inclination of 1 in 6. Assume that the frictional resistance on the incline is the same as that on the level road.

(10^4 watts; 2960 N)

2.29 An electric motor is used to lift an elevator and its load (total 1500 kg) to a height of 20 m. The time taken for the job is 20 s. The efficiency of the motor is 75 per cent. What is the work done? What is the rate at which work is done? At what rate is the energy supplied to the motor?

(2.94×10^5 j; 1.47×10^4 watts; 1.96×10^4 watts)

UNIT 3

Circular Motion

When we travel by train, often we find the train going along an arc of a circle. Similarly, a road track also takes the shape of an arc of a circle here and there. Therefore, when we go on a bicycle or in a car we keep to a circular path now and then. We find athletes running in circular tracks. The moon goes round the earth in an approximately circular orbit. It is worthwhile studying motion in a circle. We shall confine our study to the motion of a body in a circular orbit with uniform speed.

3.1 Angular Velocity

If a body is rotating about an axis and has turned through an angle θ in time t, then $\frac{\theta}{t}$ gives the angular velocity ω (omega) of the rotating body about that axis.

NOTE: θ/t gives the average value, if the angular velocity is variable. If it is not variable, θ/t gives the constant value of the angular velocity.

If θ is in radians and t in seconds, ω will be in radians per second (rad. s^{-1}).

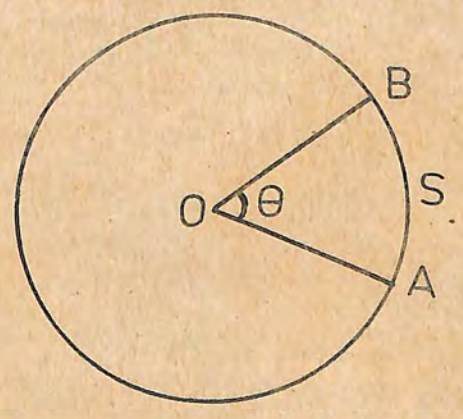


Fig. 3.1 Angular velocity: $\omega = \theta/t$.

Let a particle move with a uniform speed in a circle of a radius r (Fig. 3.1). Its angular velocity $\omega = \frac{\theta}{t}$ ... (3.1)

θ being the angle described at the centre 0 in time t. But the angle θ subtended at the centre by the arc AB = Arc/radius

$\therefore \theta = \frac{s}{r}$ where s is the length of the arc AB covered by the particle in time t.

$$\therefore \omega = \frac{\frac{s}{r}}{t} = \frac{s}{rt} = \frac{1}{r}\left(\frac{s}{t}\right)$$

or $\omega = \frac{1}{r}v$, where v is the speed of the body.

$\therefore v = r\omega.$ ... (3.2)

The dimensions of the angular velocity are $\frac{\frac{L}{T}}{L} = T^{-1}$.

3.2 Uniform Circular Motion

Consider a particle of mass m moving along a circle of a radius r with uniform angular velocity ω. Then the speed of the particle will be uniform and equal to $r\omega$. Let A be the position of the particle at any instant of time and B its position after a short time interval Δt. The initial velocity is given by the vector V_1 and the velocity after time Δt by the vector V_2. The magnitude v of the velocity is the same in both cases, but the direction has changed. This is a case of change in velocity due to change in direction.

Since the velocity of the particle is changing, it has acceleration which can be obtained by finding the rate of change of velocity.

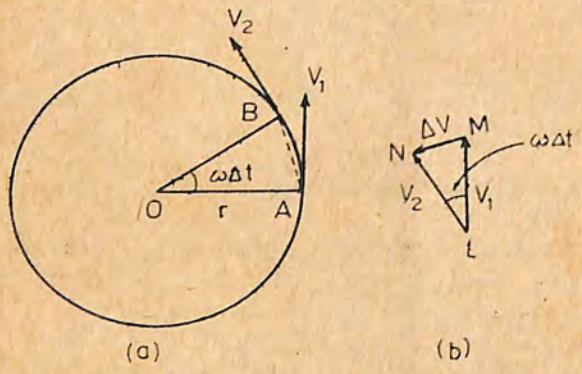


Fig. 3.2 Uniform circular motion

In Fig. 3.2b, LM and LN represent the two velocity vectors V_1 and V_2, respectively. The final velocity V_2 is the vector sum of the initial velocity V_1 and the change in velocity ΔV. MN denotes the change in velocity.

Δ^s LMN and OAB are similar, since $\angle$ BOA $= \angle$ NLM and the ratio of the sides OA to OB is equal to the ratio of the sides LM to LN, each ratio being 1.

$\therefore \frac{MN}{AB} = \frac{LM}{OA}$. As $\Delta t \to 0$, the angle $\omega \Delta t$ also tends to zero and side AB approximates to the arc AB. On replacing AB by $v \Delta t$ in the above equation, we have

$$\frac{\Delta v}{v \cdot \Delta t} = \frac{v}{r} \text{ or } \frac{\Delta v}{\Delta t} = \frac{v^2}{r}.$$

When Δt tends to zero, we get the instantaneous acceleration.

$\therefore$ The acceleration is equal to $\frac{v^2}{r}$. Since v is constant, the average acceleration and the instantaneous acceleration are the same. The acceleration is also equal to $r\omega^2$ since $v = r\omega$.

$$\therefore \text{Acceleration } |a| = \frac{v^2}{r} = r\omega^2 \quad \ldots (3.3)$$

Acceleration is a vector and its direction has to be determined. When the value of the angle $\omega \Delta t$ is reduced, the side LN gets closer to LM and in the limit, when $\Delta t \to 0$, V_2 and V_1 become almost parallel to one another. Then MN or ΔV will be perpendicular to the velocity vector. Hence, the acceleration is perpendicular to the velocity vector. Since the direction of the velocity at any point on the circle is along the tangent to the circle, at the point, the acceleration will be along the radius through that point directed towards the centre 0. If a body is to move along the circumference of a circle, it should have acceleration directed towards the centre of the circle. This is called *centripetal acceleration*. The force F towards the centre is called the *centripetal force* and is given by

$$F = \frac{mv^2}{r} = mr\omega^2. \quad \ldots (3.4)$$

Centripetal in Greek means 'seeking centre'.

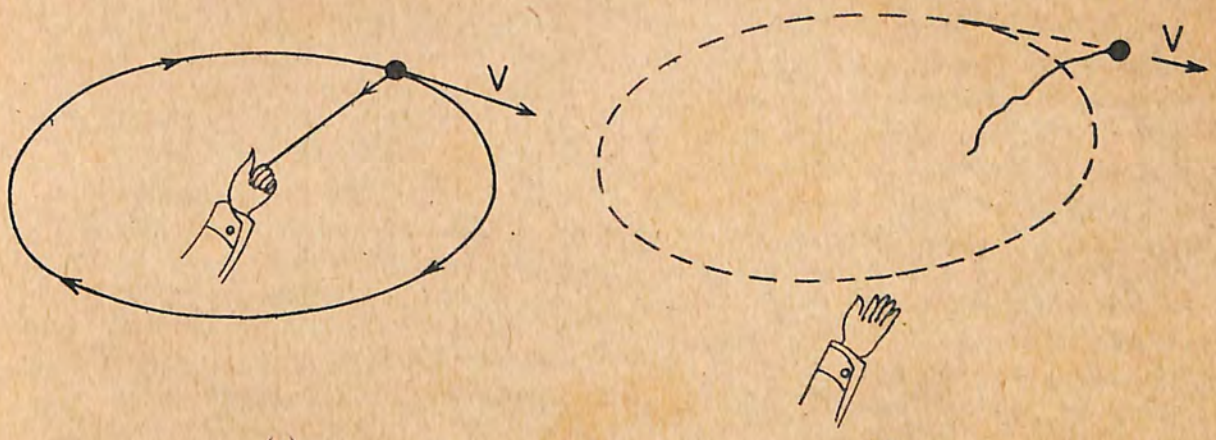
When a ball at the end of a string is whirled in a circle (Fig. 3.3 a), the hand exerts a pull and provides the centripetal force necessary to keep the ball in a circular orbit. If the string is released, the ball will not move along the radius because the moment the string is released the centripetal force on the body ceases to act. But is has a velocity v along the tangent to the circle. Hence, the body flies off tangentially when the string is released (Fig. 3.3b).

EXAMPLE 3.1

A particle of mass $\frac{1}{4}$ kg is to be whirled in a circle of a radius 1 metre with a constant speed of 4 ms^{-1}. What is the required centripetal force?

The centripetal force $F = \frac{mv^2}{r}$.

In this example, $m = \frac{1}{4}$ kg, $r = 1$ metre, $v = 4\ ms^{-1}$.



(a) (b)

Fig. 3.3 (a) A ball whirled in a circle. (b) When the string is released, the ball flies off tangentially.

$$\therefore F = \frac{1}{4} \times \frac{16}{1} = 4N.$$

EXAMPLE 3.2

An automobile of mass 1200 kg has to make a circular turn of a radius 40 metres. If it is moving with a speed of 10 ms^{-1}, what is the centripetal force on the automobile?

The centripetal force $F = \frac{mv^2}{r}$.

In this example, $m = 1200$ kg, $r = 40$ m, $v = 10\ ms^{-1}$.

$$\therefore F = \frac{1200 \times 10 \times 10}{40} = 3000N.$$

3.3 Banking of the Curve

If it is a level road on which an automobile is going in a curved path, the centripetal force will be provided by the friction between the wheels and the road (Fig. 3.4). In such a case, the tyres will soon wear out. We see, in practice, that when a road takes a curve, it is given an inclination sloping upwards towards the outer circumference. This reduces the wearing out of the tyres because the centripetal force will, in this case, be provided by the horizontal component of the normal reaction between the wheels and the road. This system of providing an inclination for the road on curve is called *banking of the curve*.

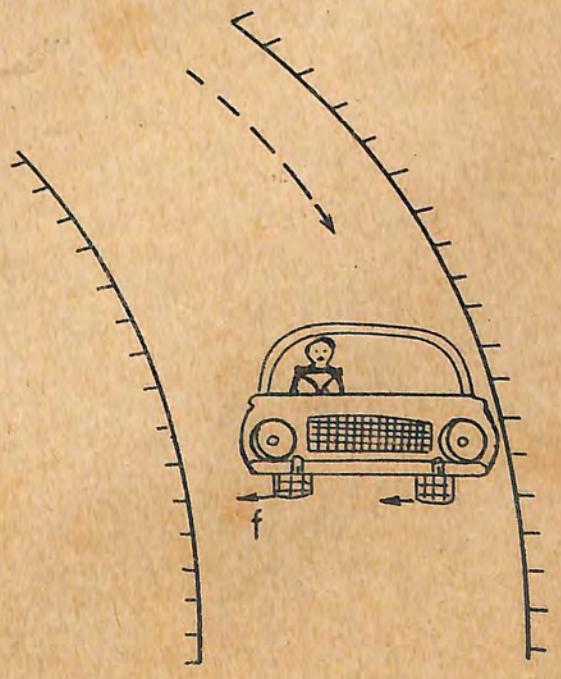


Fig. 3.4 An automobile going in a curved path. The force of friction F between the wheels and the road provides the centripetal force.

Let a car be going on a 'banked' curved road. Fig. 3.5 shows a vertical section. AB is the width of the road. It is inclined at an angle θ to the horizontal so that the normal reaction N between the wheels and the road has a horizontal component. This component of N provides the necessary centripetal force for the car to go in a circular track. In this case, there is no need for friction (between the wheels and the road) to provide the necessary centripetal force. In other words, the wear and tear of the wheels is eliminated.

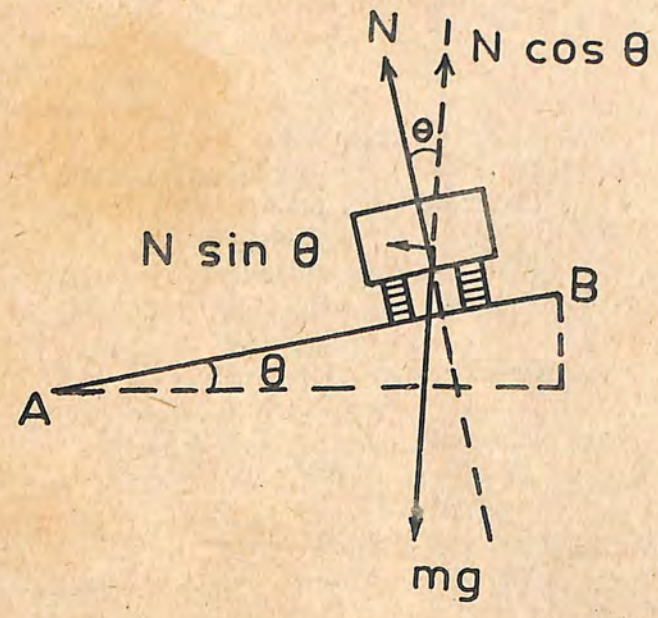


Fig. 3.5 A car going on a banked curved road (vertical section).

Resolving the normal reaction N, between the wheels and the road, along the vertical and the horizontal directions, we have:

$$N\cos\theta = mg \text{ (since there is no movement along the vertical).}$$

$$N\sin\theta = \frac{mv^2}{r} \text{ (centripetal force)}$$

where v is the speed of the car, r is the radius of the curved track and m is the mass of the car.

Taking the ratio, we get

$$\tan\theta = \frac{v^2}{rg}. \qquad \ldots (3.5)$$

The speed corresponding to the inclination of the road is generally indicated on a signboard before the curve begins. A similar thing ('banking of the curve') is done in the case of curved railway tracks also. The outer rail is elevated with respect to the inner rail in the curved part of the track.

EXAMPLE 3.3

A spherical cage is erected for performing a motorcycle feat of 'going in a vertical circle'. The radius of the spherical cage is 5 m. What is the minimum speed with which he has to perform the feat so that he may not fall? Assume the value of g to be 9.8 ms^{-2}.

The weight of the motor cycle and the cylist together acts vertically downwards. Let it be mg. If he is to keep to the vertical circle without falling, the requirement is that the centripetal force should be greater than mg, i.e., $\frac{mv^2}{r} > mg$.

The minimum speed is given by the condition that $\frac{mv^2}{r} = mg$, or $v^2 = rg$. In this example, $v^2 = 5 \times 9.8$.

$$\therefore v = \sqrt{49.0}$$
$$= 7.0\ ms^{-1}.$$

In kilometres per hour, the speed v = 25.2 km/hr, m, the mass of the cyclist and of the motor cycle, cancels out of the equation. Generally, to allow for the safety factor, he goes at a considerable higher speed.

EXAMPLE 3.4

The total mass of a motor cycle and the rider is 120 kg. The rider has to negotiate a curve of a radius 60 m at a speed of 15 ms^{-1}.

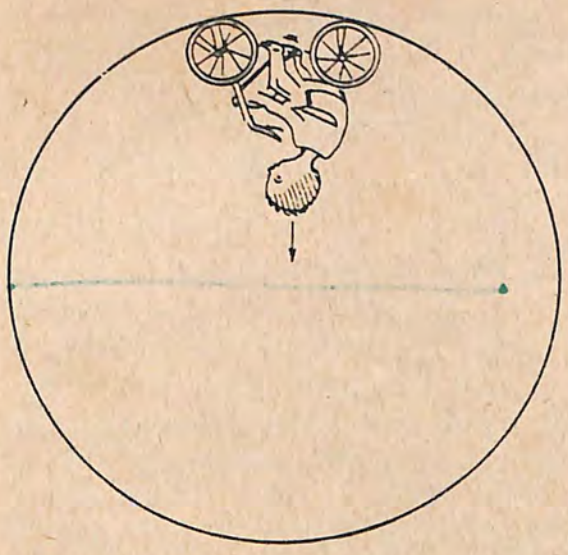


Fig. 3.6 A motor cyclist going in a vertical circle

If the coefficient of friction is 0.5, will the rider be able to go around the curve successfully? At what angle should he lean over so as to avoid a spill? Acceleration due to gravity is 9.8 ms^{-2}.

The required centripetal force is:

$$\frac{mv^2}{r} = \frac{120 \times 15 \times 15}{60}$$

$$= 450 \text{ N}.$$

The total weight of the motor cycle and the rider is:

$$120 \times 9.8 \text{ N}$$

The maximum force of friction is:

$$\mu(\text{Normal force}) = 0.5 \times 120 \times 9.8$$

$$= 588 \text{ N}.$$

Since the required centripetal force is less than the maximum possible force of friction, the cyclist will be able to negotiate the curve.

If θ is the angle at which he should lean from the vertical,

$$\tan\theta = \frac{450}{120 \times 9.8} = 0.383$$

$$\therefore \theta = 21°.$$

3.4 Planetary Motion and Kepler's Laws

Introduction

Of the early astronomers, the one who made a detailed study of the motion of the planets and advanced a theory with regard to their relative situation and movements was Ptolemy, a Greek astronomer of the second century A.D. The main contention of his theory was that the earth constituted the centre of the universe around which the moon, other planets, the sun and even the stars revolved. This is called the *Geocentric theory*. The orbits proposed for the revolution of these celestial bodies were very complex, and the theory was complicated. It was believed to be true till the time of Copernicus, who, in the sixteenth century, propounded the heliocentric theory, with the sun at the centre and the planets revolving around the sun in their respective orbits. There was a lot of controversy for a long time between the protagonists of these two theories. We will be doing a disservice to the great Indian mathematician and astronomer Aryabhata of the fifth century A.D. if we fail to mention his remarkable achievement at this stage. From his astronomical observations, he stated that the earth moves about its axis*. Perhaps due to lack of communication between scientists of the east and those of the west his findings not reach the natural philosophers of the west. Even otherwise the Christian church, which dominated over every sphere of activity in Europe, was very dogmatic with regard to its ideas of the universe and opposed any theory which did

* To be precise, Aryabhata's universe was geocentric but not geostatic.

not give a central place of importance to the earth. Even Copernicus, who propounded the heliocentric theory, could not boldly assert it for fear of the church but gave it as a mathematical convenience for understanding the motion of the planets.

Tycho Brahe, a great astronomer of the sixteenth century (1546–1601), made observations of the movements of the planets over many years. Kepler (1571–1630) was his assistant for some time before the death of Tycho Brahe. He made a careful analysis of the data collected by Tycho Brahe. Kepler noticed certain regularities with regard to the motion of the planets and formulated three laws of planetary motion which go by his name.

3.5 Kepler's Laws

Kepler's laws of planetary motion are as follows:

(i) Each planet revolves around the sun in an elliptical orbit, with the sun at one focus of the ellipse (the law of orbits).

(ii) The radius vector from the sun to the planet sweeps out equal areas in equal times (the law of areas).

(iii) The square of the period of any planet about the sun is proportional to the cube of the planet's mean distance from the sun (the law of periods). In other words, T^2/r^3 is the same for all the planets where T is the period of the planet i.e., the time taken by the planet for one complete revolution around the sun and r is the mean distance of the planet from the sun.

The speed of a planet in its orbit varies but the variation is such that the radius vector sweeps out equal areas in equal times. The acceleration of the planet and hence the gravitational force is always directed towards the sun.

Kepler's laws, as already stated, were the result of a careful analysis of the astronomical data that were available to him from Tycho Brahe's observations. These laws were empirical. There was no theoretical background or interpretation.

3.6 Newton's Universal Law of Gravitation

Aristotle, a Greek philosopher, believed that the motion of the celestial bodies is governed by laws different from those governing the motion of bodies on the surface of the earth. Till the seventeenth century, people accepted as a matter of faith that all bodies on the surface of the earth have the inherent property of falling to the earth when released from above which needed no further explanation. It only occurred to Newton that the falling of bodies towards the earth is due to a force of attraction between the bodies and the earth. He thought that the force of gravitation that attracts the apple to the earth must be also responsible to keep the moon going round the earth by providing the necessary centripetal force. He estimated the centripetal acceleration of the moon from the knowledge of its period of revolution and the radius of its orbit. He got a low value of about 0.00267 ms^{-2}. To explain the difference between the acceleration due to gravity on the surface of the earth and the centripetal acceleration of the moon in its orbit, he assumed that the force of attraction is inversely proportional to the square of the distance of the body from the earth. Another far-reaching assumption he has made is to consider the mass of the earth to be concentrated at its centre. This assumption was justified by him mathematically by using calculus which was invented by him.

The distance of the falling apple from the centre of the earth is the radius of the earth, which is equal to 6.4×10^3 km; while the distance of the moon from the centre of the earth is 3.85×10^5 km. (radius of the orbit of the moon round the earth). He assumed that,

$$\frac{\text{Acceleration of the apple}}{\text{Acceleration of the moon}} = \frac{(\text{Radius of the moon's orbit})^2}{(\text{Distance of the apple from the centre of the earth})^2}$$

The left-hand side expression is equal to $\frac{9.8}{0.00267} = 3675$, assuming the acceleration due to gravity on the surface of the earth to be 9.8 ms^{-2}. The right-hand side expression is equal to

$$\frac{(3.85 \times 10^5)^2}{(6.4 \times 10^3)^2} = 3619.$$

The two ratios which are in good agreement give evidence in favour of Newton's ideas.

Since forces on bodies depended on their mass (F = ma), he concluded that the gravitational force of the earth on the moon should depend upon the mass of the moon. Similarly, the force of the moon on the earth should depend upon the mass of the earth. Hence, he concluded that the gravitational force between the moon and the earth should depend upon both the mass of the earth and the moon, besides being proportional inversely to the square of the distance between them. Similarly, the force of attraction on the apple should depend upon both the *mass* of the earth and the apple and inversely proportional to the square of the distance between them. He generalized these conclusions and thought that between any two bodies in this universe there exists a force of attraction directly proportional to the mass of the two bodies and inversely proportional to the square of the distance between them. He could not put his ideas to experimental test by measuring attraction between two bodies in the laboratory for the simple reason that these forces must be small and that other forces like friction, etc., predominated over these forces.

He tried to put his ideas to test in the case of planetary motion and check up whether they lead to the results in conformity with Kepler's laws. He assumed that each planet is attracted towards the sun with a force directly proportional to the mass of the planet and that of the sun and inversely proportional to the square of the planet's distance from the sun. On the basis of his assumptions, Newton arrived at Kepler's law of periods*,

---

* For a planet, if the orbit is circular, we have the relation

$$T = \frac{2\pi r}{v}$$

where T is the period of revolution, r is the radius of the orbit and v is the speed of revolution.

$$\therefore T^2 = \frac{4\pi^2 r^2}{v^2} \text{ and } \frac{T^2}{r^3} = \frac{4\pi^2}{rv^2} \quad \text{—} \quad (1)$$

If a is the acceleration of the planet towards the sun

$$a = \frac{v^2}{r} \text{ (for a circular orbit)}$$

According to Newton's idea, the force of attraction of any planet to the sun is inversely proportional to the square of the distance between them so that

a should be $= \frac{K}{r^2}$ where K is a constant.

Equating the two expressions for a, we have:

$$\frac{v^2}{r} = \frac{K}{r^2} \text{ or } rv^2 = K$$

Putting K for rv^2 in the equation (1), we get

$$\frac{T^2}{r^3} = \frac{4\pi^2}{K} = \text{constant.}$$

If K depended on the mass of the planet, $\frac{T^2}{r^3}$ will not be the same for all planets.

What we have deduced is applicable to elliptical orbits also, only if we define r to be the semi-major axis of the ellipse.

namely, $\frac{T^2}{r^3}$ is the same for all planets.

His inverse square law of gravitational force between the sun and any planet shows that the orbit of any planet must be an ellipse with the sun at one focus.

Newton's universal law of gravitation is stated as follows:

> Every body in the universe attracts every other body with a force directly proportional to the product of their mass and inversely proportional to the square of the distance separating them.

In mathematical notation, it may be be written as

$$F = G \frac{m_1 m_2}{r^2} \quad \ldots (3.6)$$

where m_1 and m_2 are the masses of the two bodies and r is the distance separating them. G is a constant called the *universal gravitation constant*.

The fact that Newton could explain Kepler's laws of planetary motion on the basis of his ideas of motion and gravitation gave support to his law of gravitation. The grand feature of his triumph was that he could explain the motion of the planets as well as the motion of the bodies falling near the surface of the earth by the same laws. This gave a death blow to the earlier belief that the motion of the celestial bodies was governed by laws different from those governing the motion of bodies on the earth (terrestrial bodies). Further, the heliocentric theory of Copernicus was confirmed by the work of Kepler and Newton. It provided an inertial reference frame, fixed with respect to the sun.

3.7 Universal Gravitation Constant

The value of the gravitation constant G does not depend on the nature of the bodies or the properties of the intervening medium, if any. Hence, the universality of the law of gravitation.

The dimensions of G are $L^3 M^{-1} T^{-2}$.

The constant of gravitation G can be determined, provided one can measure the gravitational force f between two bodies of the known mass m_1 and m_2 separated by a distance r. The gravitational force between any two objects of laboratory size is very small, and the measurement of forces of that low order was not known for a long time after Newton. Henry Cavendish succeeded in determining the value of G in the year 1798. Improved techniques were adopted by those who made subsequent determinations. The value of G accepted now is 6.67×10^{-11} Nm^2/kg^2.

EXAMPLE 3.5

Calculate the mass of the earth, given that

the gravitation constant $G = 6.67 \times 10^{-11}$ Nm^2/kg^2

the radius of the earth $R = 6.38 \times 10^6$ m

and the acceleration due to gravity (g) = $9.8\ ms^{-2}$.

We know that the gravitation force of attraction of the earth on body of mass m near the surface of the earth is given by

$$mg = G \frac{Mm}{R^2},$$

where M is the mass of the earth

$$\text{or } M = \frac{gR^2}{G}$$

$$= \frac{9.8\,(6.38 \times 10^6)^2}{6.67 \times 10^{-11}}$$

$$= 5.98 \times 10^{24} \text{ kg}.$$

3.8 Gravitational Field

In the region surrounding a magnet exists a magnetic field. Likewise, an electric field exists around an electric charge. If a charge is brought into this field, it experiences a force. Similarly there is a gravitational field surrounding the earth and if a body is

brought into this field, it experiences a force towards the centre of the earth. On the analogy of magnetic and electric field strengths, we can also define gravitational field strength.

The gravitational force F due to the earth on a body of mass m situated at a distance x from the centre of the earth (x being $>$ R, the radius of the earth) is given by:

$$F = \frac{GMm}{x^2} \qquad \ldots (3.7)$$

where M is the mass of the earth.

If m = 1, the force is given by $\frac{GM}{x^2}$. This value of the force on a unit mass at a point at a distance x from the centre of the earth is called the *gravitational field strength due to the earth at that point*. At any point in the field, the force on a unit mass gives the value of the field strength at that point.

If k is the gravitational field strength at a point, then the force F on a body of mass m situated at that point is given by

$$F = km. \qquad \ldots (3.8)$$

It may be noted that the gravitational field strength at a point near the surface of the earth is equal to g.

3.9 Gravitational Potential Energy

Potential energy is a mutual property of a pair of bodies resulting from their proximity to one another and the fact that they interact with one another. When the relative distance between two bodies is altered work is done either by the gravitational force of attraction between the two bodies or against it. In either case, there is a change in the gravitational potential energy of the system.

The gravitational potential energy (denoted by ϕ) of one body with respect to another is zero when the distance between them is infinity. The gravitational force F of the earth on a body of mass m situated at a distance x from the centre of the earth (x being greater than R, the radius of the earth) is given by the equation (3.7).

The work done in shifting the mass towards the earth through an infinitesimal distance dx is equal to F dx. If the mass is brought from infinity to the surface of the earth, the total work done is obtained by integrating F. dx from ∞ to R :

$$= \int_{\infty}^{R} \frac{GMm}{x^2}\, dx$$

$$= \left[-\frac{GMm}{x} \right]_{\infty}^{R}$$

$$= -\frac{GMm}{R}.$$

The $-$ ve sign indicates that work is done by the gravitational force of attraction; hence, the mutual energy is expended so that it is reduced by this amount. Originally, when the body was at infinity, its potential energy with respect to the earth was zero. The gravitational potential energy ϕ of a body of m near the surface of the earth is, therefore, given by

$$\phi = -\frac{GMm}{R}. \qquad \ldots (3.9)$$

3.10 Earth's Satellite

The moon is the natural satellite of the earth. It goes round the earth once in 27.3 days in a nearly circular orbit of mean radius 3.8×10^5 km. In 1956, man for the first time put in orbit round the earth an artificial satellite. Since then, many artificial satellites have been put in orbit a few hundred kilometres above the surface of the earth to study various phenomena in the outer regions of the earth's atmosphere.

It may be mentioned here that India has put in orbit on 19 April 1975 an earth

satellite 'Aryabhata' (Fig. 3.7) named after the celebrated Indian astronomer Aryabhata of the fifth century A.D. Much work is going on in India to launch some more satellites in future.

3.11 Escape Velocity

All slow-moving projectiles fall back to the earth. But rockets used for launching earth satellites have to overcome the earth's gravitational attraction. What is the velocity with which a rocket must be fired in order to just escape completely from the earth's gravitational attraction? This value of velocity is called the *escape velocity*.

Let the mass of the earth be M and that of the rocket m. After the rocket is fired and all the fuel is spent, the height attained above the earth is not much compared with the radius R of the earth. So, we can assume that the distance of the rocket from the centre of the earth is R. Let the velocity of the rocket at this stage be V_R.

At this stage, the gravitational P.E. of the system from the equation (3.9) is

$$-G\frac{M\,m}{R}$$

and the K.E. $\frac{1}{2}m v_R^2$.

Suppose, the rocket has ultimately reached

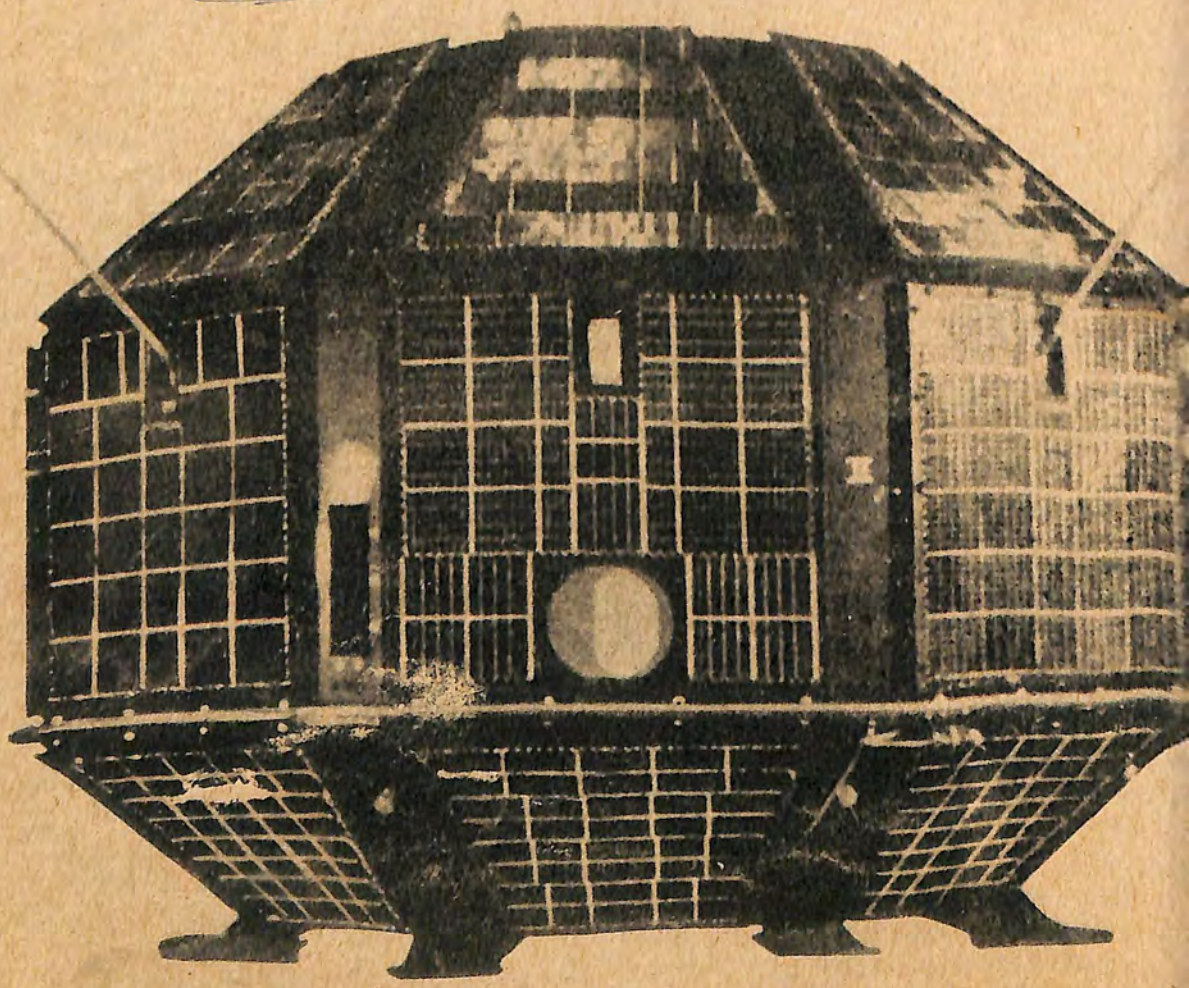


Fig. 3.7 'Aryabhatta', an earth satellite.

an infinite distance where the gravitational potential energy is zero. Then, the kinetic energy is $\frac{1}{2}mv^2$ where v is the velocity of the rocket at infinite distance. From the law of conservation of energy, we have

$$\frac{1}{2}m v^2 = \frac{1}{2}m v_R^2 - G\frac{M m}{R}.$$

The left-hand side expression, namely, the final K.E. must be positive if the rocket is to escape the attraction of the earth.

$$\therefore \frac{1}{2}m v_R^2 \text{ must be } > G\frac{M m}{R}$$

that is, v_R must be $> \sqrt{\frac{2GM}{R}}$.

We know that

$G = 6.67 \times 10^{-11}\ Nm^2/kg^2$

$M = 5.98 \times 10^{24}$ kg

and $R = 6.4 \times 10^6$ m.

$\therefore V_R$ must be $>$

$$\sqrt{\frac{2 \times 6.67 \times 10^{-11} \times 5.98 \times 10^{24}}{6.4 \times 10^6}}$$

$\therefore V_R$ must be $> 11.160\ m.s^{-1}$

or $11.16\ km\ s^{-1}$

or V_R must be > 40000

or $4 \times 10^4\ km\ h^{-1}$.

$\therefore$ The rocket must be fired with a velocity greater than 4×10^4 km/h to escape from the gravitational attraction of the earth.

NOTE: As the rocket rises, the P.E. increases from $-G\frac{M m}{R}$ and finally attains the value zero at infinity. This increase is obtained at the expense of the K.E. of the rocket.

3.12 Orbital Velocity

Artificial satellites are put in orbit at a height of a few hundred kilometres above the surface of the earth. At this height, the air is in a very highly rarefied state. A satellite is released horizontally with a very high velocity so that it remains moving in a nearly circular orbit. Since the air resistance is very very small, no work is done to keep the satellite moving in its orbit and as such no fuel is used by it.

Let us examine the question: 'Is there any prescribed value for the velocity of the satellite to continue to move in its orbit? We have seen that when a particle moves along the circumference of a circle with uniform speed, there is an acceleration on the particle towards the centre of the circular path. Consequently, there is the centripetal force. As long as this centripetal force is provided, the particle remains in this circular orbit.

Let m be the mass of a man-made satellite and h the height above the ground where it is to orbit. The gravitational force F on the satellite towards the centre of the earth is given by the equation (3.7),

$$\therefore F = G\frac{M m}{(R + h)^2}.$$

where M is the mass of the earth, R the radius of the earth and G the universal constant of gravitation. If a_s is the consequent acceleration of the satellite towards the centre of the earth, we have

$$a_s = \frac{F}{m} = \frac{GM}{(R + h)^2}.$$

If v is the velocity which the satellite should have in order to continue in a circular orbit of a radius $(R + h)$, the necessary centripetal acceleration a is given by

$$a = \frac{v^2}{(R + h)}.$$

The acceleration a has to be provided by the gravitational force. Hence,

$$a = a_s.$$

$$\therefore \frac{v^2}{(R + h)} = \frac{GM}{(R + h)^2}$$

or

$$v = \sqrt{\frac{GM}{(R + h)}}$$

EXAMPLE 3.6

Assuming that a satellite is to orbit 330 km above the surface of the earth, calculate the necessary speed with which it should be put in orbit? Also, determine the period of revolution of the satellite, given that

Radius of the earth R = 6,370 km.

Mass of the earth, M = 5.98×10^{24} kg.

G = 6.67×10^{-11} Nm^2/kg^2.

Since, h = 330 km, we have (R + h)

= 6,700 km

= 6.7×10^6 m.

Applying the above formula,

$$v = \sqrt{\frac{6.67 \times 10^{-11} \times 5.98 \times 10^{24}}{6.7 \times 10^6}}$$

$$= 7.716 \times 10^3 \text{ ms}^{-1}$$

$$= 2.8 \times 10^4 \text{ km h}^{-1}.$$

Knowing the distance the satellite covers for one complete revolution (i.e., the circumference of the orbit), we can find the period of the satellite which is the time for orbiting once around the earth.

The circumference of the orbit

$$= 2\pi(R + h)$$

$$= 2\pi(6.7 \times 10^6) \text{ m}.$$

Velocity = 7.716×10^3 ms^{-1}.

∴ Period of revolution of the satellite around the earth

$$= \frac{2\pi \times 6.7 \times 10^6}{7.716 \times 10^3 \times 60} \text{ minutes}$$

= 91 minutes.

3.13 Launching of a Satellite

Artificial earth satellites are multi-stage rockets (Fig. 3.8). At start, the rocket must be given sufficient velocity to rise to the required altitude. When it has ascended the required height, the satillite (last stage) is discharged horizontally (Fig 3.8) with a sufficient velocity (orbital velocity) to keep it in its orbit. If the satellite is discharged with a greater velocity, it would move outward to a larger orbit or escape completely. If it is discharged with a lesser velocity, it would fall back to the earth (Fig. 3.9). The velocity being the calculated value for a circular orbit, if the satellite is not launched horizontally it moves in an elliptical orbit.

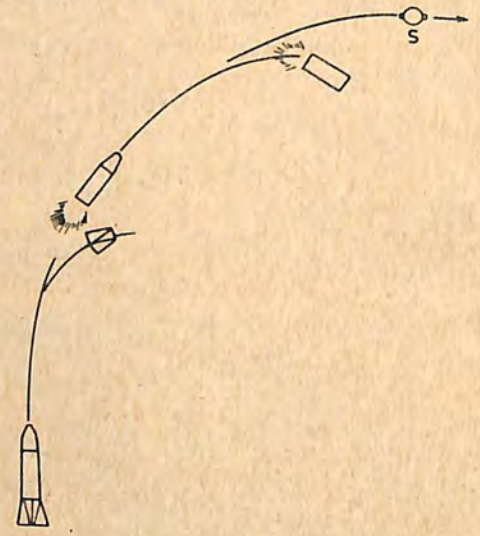


Fig. 3.8 A multi-stage rocket: S—satellite.

3.14 Rigid Body

A rigid body ideally is one whose constituent particles retain their relative positions even in motion. This implies that the body cannot be deformed. If the body undergoes some displacement, every particle in it suffers the same displacement. If the body rotates through a certain angle about an axis, every particle in it rotates through

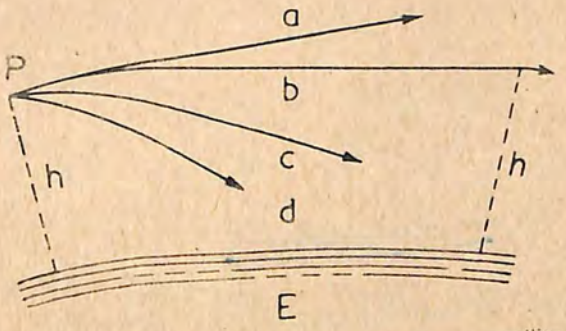


Fig. 3.9 The launching of a satellite. If the satellite is discharged horizontally with sufficient velocity, it moves in a circular orbit. E—earth; a—1.5 km/sec; b—7.75 km/sec; c—4 km/sec; d—3 km/sec; h—400 km.

the same angle about this axis. Fig. 3.10 shows a rigid body rotating about an axis through O perpendicular to the plane of the figure. The angle of rotation POP′ of any particle P in the rigid body is equal to the angle of rotation QOQ′ of any other particle Q in the rigid body. It may be noted that the points on the axis are stationary.

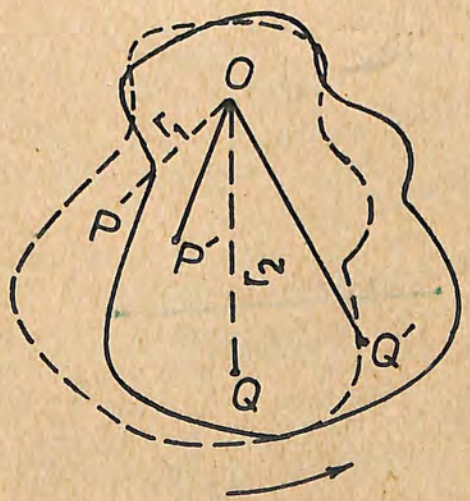


Fig. 3.10 A rigid body rotating about an axis through a point O in the body.

In everyday life, we come across several instances of rotation. Wheels, pulleys, the blades of a fan, doors revolving about a hinge and phonograph records are some examples of rigid bodies in rotational movement. The wheel of an automobile rotates about its axle and at the same time has translational motion, i.e., its position changes from moment to moment. A tennis ball moving with a spin is an example of a body having, simultaneously, rotational and translational motion. The earth is another example of a body having, simultaneously, rotational and translational motion. In some cases, the position of the axis changes with time as in the case of a precessing top tilted at an angle to the vertical. We shall confine our study only to rigid bodies rotating about a fixed axis.

It may be noted that every particle in a rigid body that is in rotation has the same value for angular velocity ω. Fig. 3.11 shows a rigid body rotating about an axis OA. Since $v = r\omega$, the farther a particle is situated from the axis, the greater is its linear velocity. P and Q are two particles at distance r_1 and r_2, respectively, from O. r_1 is less than r_2. Therefore, the linear velocity of P, namely, $r_1\omega$ is less than that of Q which is $r_2\omega$. Their linear velocities are in the ratio r_1 to r_2.

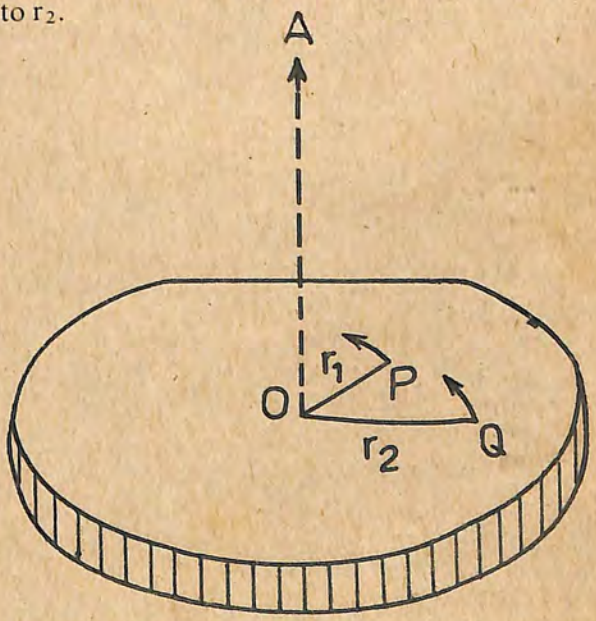


Fig. 3.11 A rigid body rotating about OA as the axis. The angular velocity of the particle P is the same as that of the particle Q. But the linear velocity of Q is greater than that of P.

EXAMPLE 3.7

What is the angular velocity of rotation of the earth about its axis? Assuming the equatorial radius of the earth to be 6,378 km, calculate the linear speed of a particle at the equator.

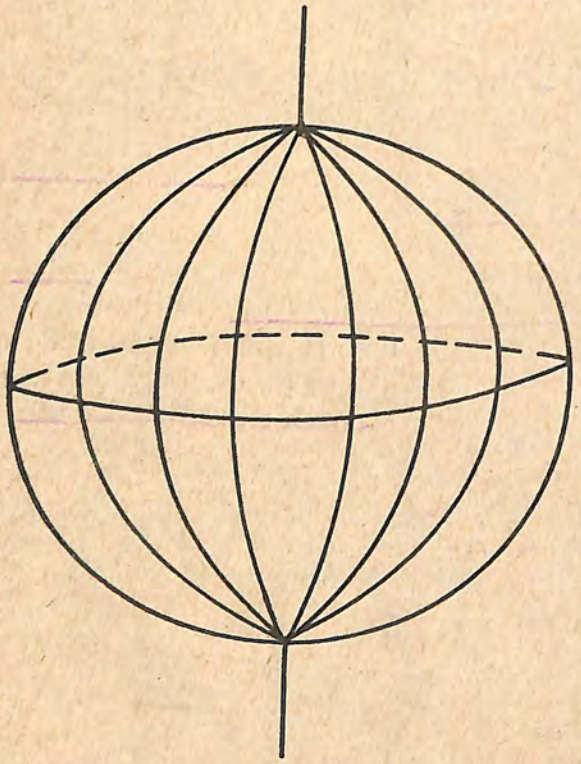


Fig. 3.12 The rotation of the earth about its axis.

The earth makes one complete rotation about its axis in 24 h.

$\therefore$ In 24 h, the earth turns through an angle 2π radians.

$\therefore$ The angular velocity ω is given by

$$\frac{\theta}{t} = \omega = \frac{2\pi}{24 \times 60 \times 60} \text{ radians per sec.}$$

The linear speed of a particle at the equator

= (equatorial radius of the earth) (angular velocity of rotation)

$$= \frac{6{,}378 \times 2\pi}{24 \times 60 \times 60}$$

$$= 0.4635 \text{ km s}^{-1}$$

$$= 1669 \text{ km h}^{-1}.$$

i.e. $v = r\omega$

3.15 Angular Acceleration

Let a particle rotating about an axis have an angular displacement θ in time t. If its angular velocity is variable, the instantaneous value of angular velocity is given by

$$\omega = \frac{d\theta}{dt} = \dot{\theta}. \qquad \ldots (3.10)$$

The angular acceleration at any instant is

$$\omega = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2} = \ddot{\theta}. \qquad \ldots (3.11)$$

Since the angular velocity ω of the particle is variable, its linear velocity $r\omega$ also is variable. Its linear acceleration a is given by

$$a = r\dot{\omega} \qquad \ldots (3.12)$$

When the angular velocity is uniform, $\dot{\omega} = 0$ so that the linear acceleration of the particle is zero.

3.16 Rotational Kinetic Energy

Fig. 3.13 shows a rigid body rotating with angular velocity ω about an axis through 0 perpendicular to the plane of the figure. P is a particle of mass m_1 at a distance r_1 from the axis. The kinetic energy of this particle is

$$= \frac{1}{2} m_1 v_1^2 = \frac{1}{2} m_1 r_1^2 \omega^2.$$

The body is made up of a large number of such particles.

$\therefore$ The total K.E. of the body

= the sum of the K.E. of all the particles composing the body.

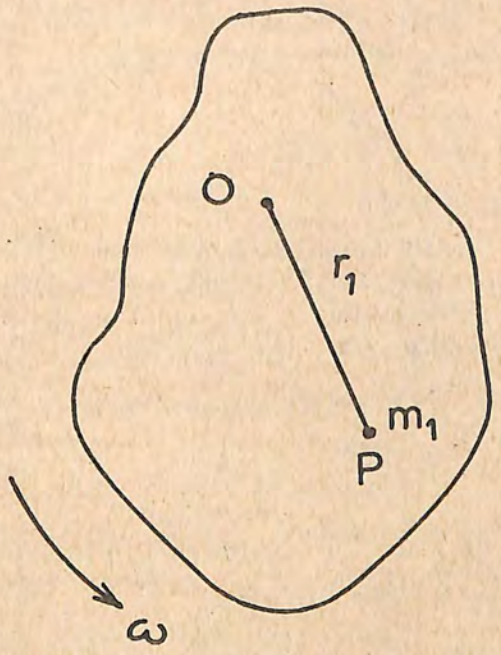


Fig. 3.13 Rotational kinetic energy.

Since every particle has the same angular velocity,

K.E. of the body

$$= \frac{1}{2} m_1 r_1^2 \omega^2 + \frac{1}{2} m_2 r_2^2 \omega^2$$

$$+ \frac{1}{2} m_3 r_3^2 \omega^2 + \ldots$$

$$= \frac{1}{2} (m_1 r_1^2 + m_2 r_2^2 + m_3 r_3^2 + \ldots) \omega^2$$

$$= \frac{1}{2} (\sum_i m_i r_i^2) \omega^2.$$

The quantity $\sum_i m_i r_i^2$ is known as the moment of inertia of the body about the axis through 0. It depends upon not only the mass of the particles of which the body is composed but also their distances from the axis. The moment of inertia is denoted by the letter I.

$$\therefore I = \sum_i m_i r_i^2. \qquad \ldots (3.13)$$

It has the same value, irrespective of the state of motion of the body. The moment of inertia of a body about an axis is constant. For the same body, the moment of inertia (M.I.) is different for different axes of rotation. Its dimensions are ML^2. The units for the moment of inertia are kg m^2 in the mks system and gm cm^2 in the egs system. Comparing the expressions for K.E. in rotational and translational motion, we find that the moment of inertia is the rotational analogue to the mass.

3.17 The Moment of Inertia of a Uniform Ring about an Axis Through its Centre 0 Perpendicular to the Plane of the Ring

Let m be the mass of a small element of a thin uniform ring (Fig. 3.14) of a radius r. The entire ring can be considered to be made up of a large number of such small elements. Every element is at the same distance from the axis through 0. Hence, the moment of inertia of the entire ring is given by

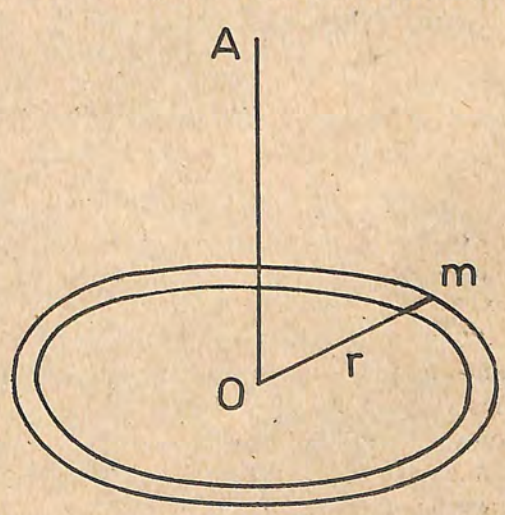


Fig. 3.14 The moment of inertia of a uniform thin ring.

$$I = m_1 r^2 + m_2 r^2 + m_3 r^2 + \ldots$$

$$= r^2 (\sum_i m_i) = M r^2. \qquad \ldots (3.14)$$

where M is the mass of the ring.

3.18 The Moment of Inertia of a Uniform Circular Disc about an Axis Through its Centre Perpendicular to the Plane of the Disc

Let m be the mass per unit area of a uniform circular disc. Out of this disc, let us consider a concentric ring of radius x and of very small thickness $\triangle x$. The mass of the ring $= (2\pi x \triangle x)\, m$.

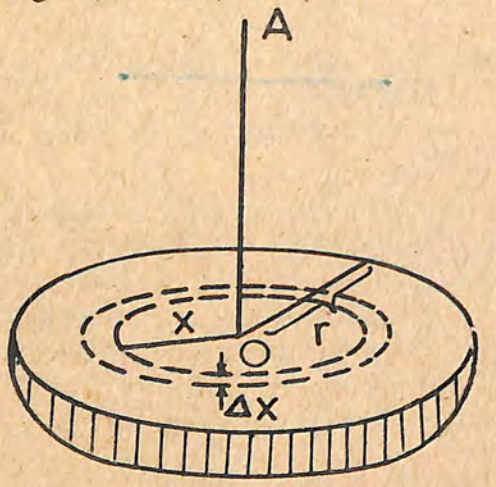


Fig. 3.15 The moment of inertia of a uniform circular disc.

The moment of inertia of the ring about an axis through the centre 0 and perpendicular to the plane of the ring is equal to $(2\pi x \triangle x)\, mx^2$.

The entire disc can be considered to be made up of a large number of such concentric rings varying in radius from zero to r. The moment of inertia I of the entire disc can be obtained by integrating the above expression between the limits 0 to r. Thus,

$$I = \int_0^r (2\pi x\, dx)\, mx^2.$$

$$= 2\pi m \int_0^r x^3\, dx.$$

$$= 2\pi m \left[\frac{x^4}{4}\right]_0^r$$

$$= \frac{2\pi mr^4}{4} = \frac{1}{2}(\pi r^2 m)\, r^2.$$

Since $\pi r^2 m$ is the entire mass M of the disc, we can write

$$I = \frac{1}{2} M r^2. \qquad \ldots (3.15)$$

EXAMPLE 3.8

A uniform circular disc of mass 0.5 kg has a radius of 10 cm. What is its moment of inertia about an axis through its centre perpendicular to the plane of the disc?

The moment of inertia of the disc

$$= \frac{1}{2} Mr^2.$$

$$= \frac{1}{2} \times 0.5 \times \left(\frac{10}{100}\right)^2$$

$$= 25 \times 10^{-4}\ \text{kg m}^2.$$

NOTE: The formula for the moment of inertia of a right cylinder about its axis is the same as that obtained above for a uniform circular disc, namely, $\frac{1}{2} Mr^2$, because a cylinder is a disc of large thickness.

Moment of inertia in a few simple cases

The formulae for moments of inertia in a few simple cases are given below:

(i) The moment of inertia of a uniform rod of mass M and length l about an axis through its centre and perpendicular to the length of the rod. $= \frac{Ml^2}{12}$

(ii) The moment of inertia of a uniform rod of mass M and length l about an axis through one end and perpendicular to the length of the rod. $= \frac{Ml^2}{3}$

(iii) The moment of inertia of a sphere of mass M and radius r about any diameter. $= \frac{2}{5} Mr^2$

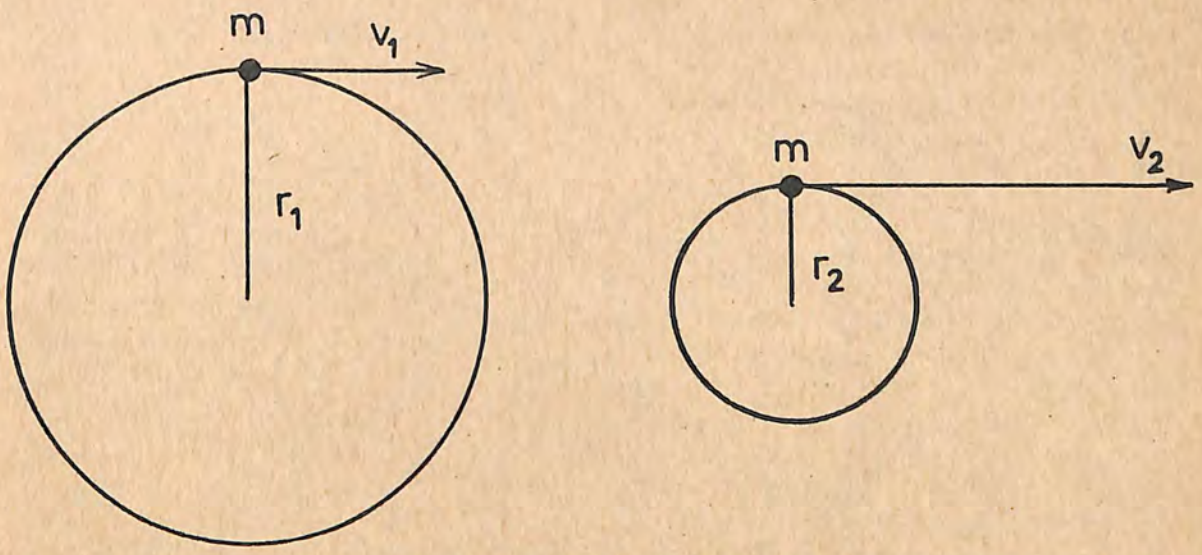


Fig. 3.16 A body whirled in a circle. When the radius of the circle is suddenly reduced, the speed of the body increases so as to keep its angular momentum constant.

3.19 Angular Momentum

Let a body of mass m be tied at the end of a string and whirled in a circle of radius r_1 with uniform speed v_1. Suppose, the radius of the circle is suddenly reduced to r_2 by holding the string at a shorter distance r_2 from the body. It will be found that the speed of the body suddenly increases. The increase in velocity will be in proportion to the decrease in the radius of the circular path, so that

$$\frac{v_2}{v_1} = \frac{r_1}{r_2} \text{ or } v_2 r_2 = v_1 r_1.$$

Multiplying each side by m, we have

$$mv_2 r_2 = mv_1 r_1.$$

mvr is referred to as the angular momentum of the particle and is denoted by L. Since

$$v = r\omega$$

$$L = m.r\omega r$$

$$\text{or } L = mr^2 \omega$$

$$\text{or } \boxed{L = I\omega.} \quad \ldots (3.16)$$

The angular momentum L is the analogue of the linear momentum mv. It is also a vector like the linear momentum. As can be seen from the above expression, the dimensions for L are $ML^2 T^{-1}$.

In the above example of a body describing a circle, the radius of the circular path was suddenly reduced and the velocity consequently increased to compensate for the reduction of r so as to maintain the angular momentum given by $L = mvr$ constant. Angular momentum is, therefore, conserved.

A person standing on a turntable with his hands outstretched (Fig. 3.17a) is holding a pair of weights, one in each hand. When the turntable is rotating, if he brings his hands inward (Fig. 3.17b), the speed of rotation will be found to increase considerably. When he draws inward the two weights, the moment of inertia of the weights is very much reduced since the value of r in the second case is much smaller than that in the first case.

It means a reduction in the value of I has brought about an increase in the value of ω. From this we infer that the change is in such a way as to maintain the angular momentum $I\omega$ constant. It may be noted that there is no external rotating effect or couple acting on

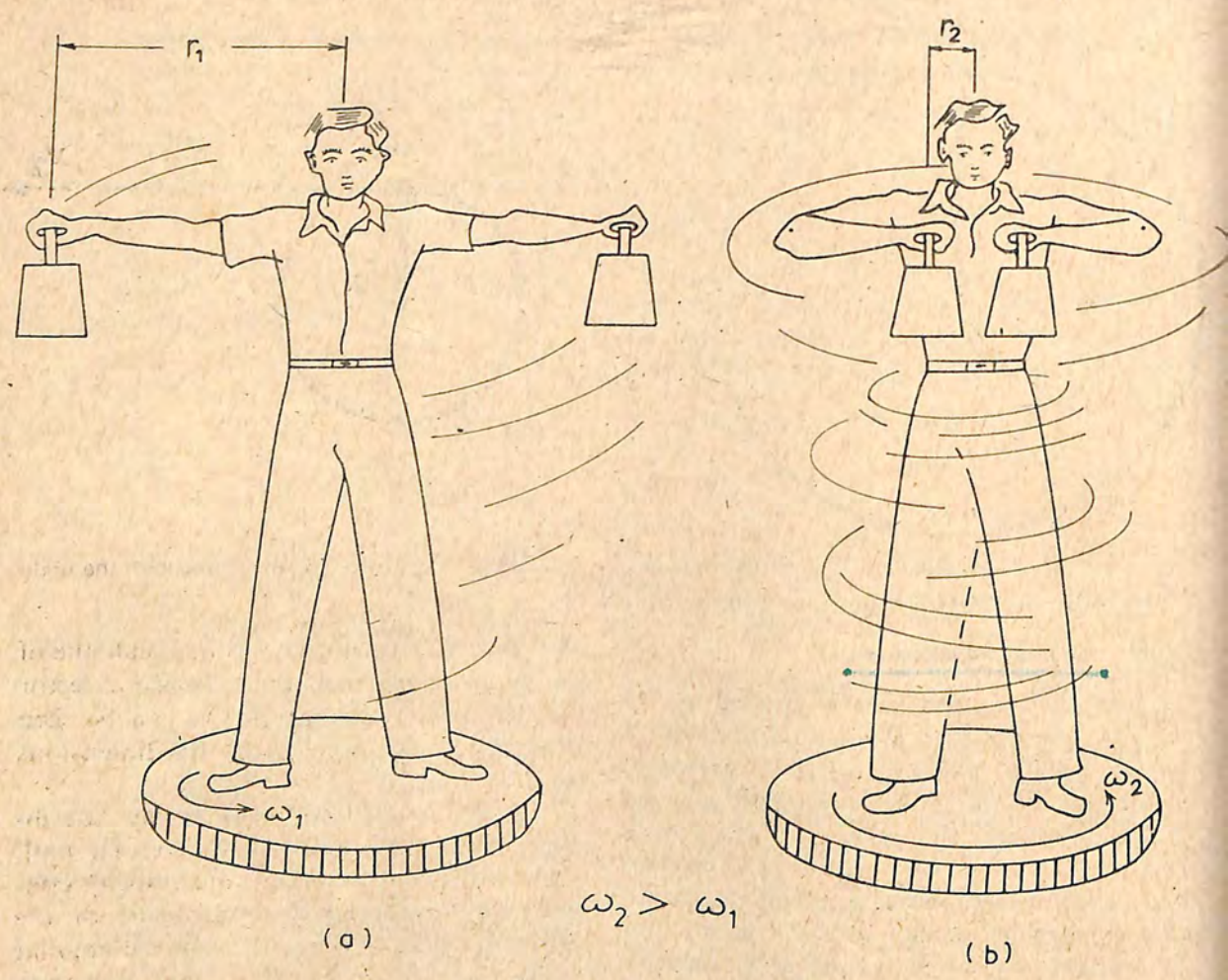


Fig. 3.17 Conservation of angular momentum. In (b), the moment of inertia is less compared with that in (a). The angular velocity in (b) has increased to maintain the angular momentum constant.

the system. (The moment of a force tending to rotate a body about an axis is called *torque*). We can conclude that in the absence of an external torque, the angular momentum of a system is conserved.

The principle of conservation of angular momentum is stated as follows:

> *When the resultant external torque acting upon a system of particles is zero, the total (vector sum) angular momentum of the system remains constant.*

It may be noted that this principle is similar to the principle of conservation of linear momentum in the case of linear motion.

3.20 Torque

Let a particle of mass m be restricted to motion in a circle of radius r (Fig. 3.18). A force **F** acting on the particle along the tangent to its path gives it an acceleration a, given by

$$a = \frac{F}{m}$$

If the angular acceleration of the particle is $\dot{\omega}$, we have

$$a = r\dot{\omega}.$$

$$F = ma = mr\dot{\omega}.$$

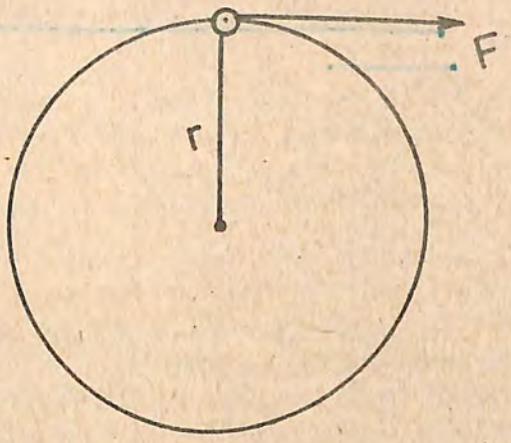


Fig. 3.18 Torque on a particle moving in a circle

Fr gives the moment of the force or torque τ (Greek letter taw) about the axis of rotation of the particle.

$$\therefore \vec{\tau} = Fr = m r^2 \dot{\omega}.$$

In this expression, mr^2 can be replaced by I, the moment of inertia of the particle about the axis of rotation, so that

$$\boxed{\vec{\tau} = I\dot{\omega}.} \qquad \ldots (3.17)$$

The same formula applies to any rotating body, provided that the torque and moment of inertia are calculated about the same axis.

We have seen in linear motion that when a force acts on a body, its momentum changes and the rate of change of linear momentum P is a measure of the force applied. Likewise, when torque acts on a rotating body, its angular momentum L changes and the rate of change of angular momentum of the body is a measure of the torque applied.

$$\text{Torque } \vec{\tau} = \frac{dL}{dt} = \frac{d(I\omega)}{dt} = \frac{I\, d\omega}{dt}$$

$$= I\dot{\omega}.$$

Therefore, torque plays the same role in rotational motion as force in translational motion. It is also a vector.

The dimensions of τ (torque) are

= (dimensions of I) (dimensions of angular acceleration)

$= ML^2T^{-2}$.

Torque is necessary to change the angular velocity of a rotating body. When a tennis ball or a football is released with a spin, it is difficult to hit it back in the desired direction because of the necessity to apply suitable torque to change the orientation of the axis of rotation of the spinning ball.

In Fig. 3.19, a wrench is used to turn a nut. If the wrench is held close to the nut as in Fig. 3.19a, the torque or the rotating

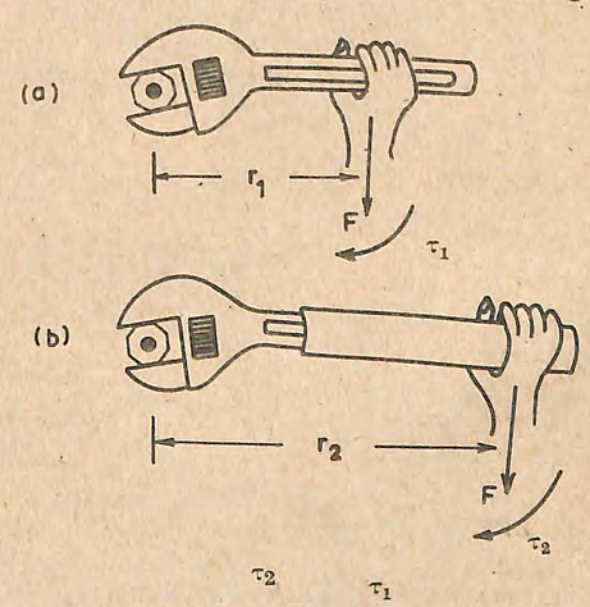


Fig. 3.19 The use of a wrench to turn a nut. Force applied being the same, the torque τ_2 in (b) is greater than the torque τ_1 in (a).

effect is low. If the wrench is held near the other end and turned, the torque will be more for the same force that is applied. Therefore, the turning effect depends upon the length of the lever arm which is the perpendicular distance from the centre of rotation of the nut on to the line of action of the force. Sometimes, a piece of pipe (Fig. 3.19b) is

slipped over the wrench to enable the application of the force at a greater distance from the nut resulting in the greater torque given by

$$\text{torque } \vec{\tau_2} = F \times r_2$$

where F is the force applied and r_2 is the perpendicular distance from the centre of rotation to the line of action of the force (Fig. 3.19b).

3.21 Work Done by Torque

Let two equal and opposite forces F, $-F$ act tangentially on the rim of a disc from the extremities of a diameter. Let the disc be free to rotate about an axis through its centre O, perpendicular to the plane of the disc. If the couple rotates the body through a

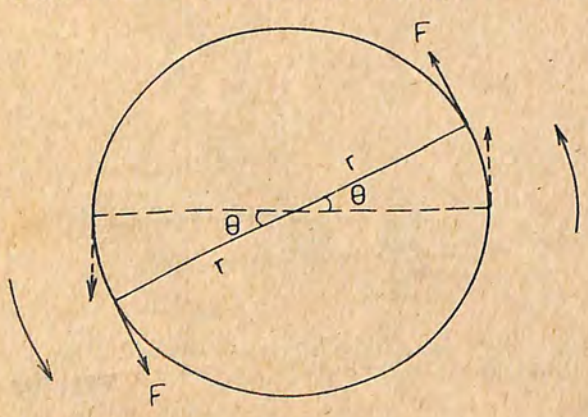


Fig. 3.20 Work done by the torque.

very small angle θ, the displacement s of each force is given by $s = r\theta$. Since θ is very small, the displacement s is almost in the same direction as F. Therefore, the work done by a single force is $Fs = Fr\theta$. The total work done by the two forces constituting the couple $= 2\,Fr\theta$. But 2 Fr or F (2r) is the moment of the couple of the torque τ.

$$\therefore \text{ work done by the torque} = \tau\theta \quad \ldots (3.18)$$

NOTE: The same formula applies even when the value of θ is large.

3.22 Torque as a Cross-product of Vectors *r* and *F*

The force F and the lever arm r are vectors. The vector moment of the force F about the point 0 is the torque $\vec{\tau}$. This can be written as the cross-product of two vectors:

$$\vec{\tau} = r \times F.$$

The magnitude of the torque is rF and its direction is perpendicular to the plane of r and F. The direction and sense of the vector product is specified by the right-hand screw rule (Article 2.7 and Fig. 2.14). QR in Fig. 2.11a represents force F and PQ represents lever arm r.

EXAMPLE 3.9

Starting from rest, the flywheel of a motor attains an angular velocity of 60 radians per second in 5*s*. What is the angular acceleration? What is the angular displacement from the start?

The initial value of angular velocity $\omega_o = 0$

Value of ω after 5 s = 60 radians/s.

The equations for rotational motion are analogous to those for linear motion.

(Final value of angular velocity) = (Initial value of angular velocity)

+ (Angular acceleration) × (time)

or $\omega = \omega_o + \dot{\omega}\,t$.

Substituting the values for ω_o, ω and t, we have

$$60 = \dot{\omega} \times 5$$

or $\dot{\omega} = 12$ radians/s^2.

On the same analogy, if θ is the angle of rotation in 5s from start, we have

$$\theta = \omega_o t + \frac{1}{2}\dot{\omega} t^2.$$

$$\therefore \theta = \frac{1}{2} \times 12 \times 5^2 = 150 \text{ radians.}$$

EXAMPLE 3.10

A uniform circular disc of mass 1 kg and of radius 0.1 m is acted upon by a torque comprising of two equal and opposite forces applied tangentially to the rim from the extremities of a diameter. The disc is free to rotate about an axis passing through its centre perpendicular to the plane of the disc. What should be the magnitude of each force so that the disc has an angular acceleration of 20 radians/s? What is the kinetic energy transferred to the disc in 4 s?

Torque $\tau = I\dot{\omega}$.

In this example, I for a disc $= \frac{Mr^2}{2}$

$$\therefore \text{torque} = \frac{Mr^2}{2}\dot{\omega} = \frac{1 \times (0.1)^2 \times 20}{2}.$$

Torque τ is also equal to $F \times 2r$ where F is the magnitude of each of the two forces.

$$\therefore F \times .2 \times 0.1 = \frac{(0.1)^2 \times 20}{2}$$

or $F = 0.5$ N.

Angular velocity ω at the end of 4s $= 20 \times 4 = 80$ radians/s

$$\therefore \text{K.E. transferred} = \frac{1}{2} I\omega^2$$

$$= \frac{1}{2} \times \frac{1 \times (0.1)^2}{2} \times 80^2$$

$= 16$ joules

EXAMPLE 3.11

A grindstone is in the form of a flat circular disc of diameter 0.5 *m* and mass 16 kg. What constant torque should be applied so that it attains an angular velocity of 120 rpm (revolutions per minute) in 8 seconds? At what rate is work done by the torque at the end of 8 *s*.

$$\text{Angular velocity } \omega \text{ after 8s} = \frac{120}{60} \times 2\pi$$

$= 4\pi$ radians/s.

$\omega = \dot{\omega}t$, since the initial angular velocity is zero.

$$\therefore 4\pi = \dot{\omega} \times 8$$

$$\text{or } \dot{\omega} = \frac{\pi}{2} \text{ radians/s}^2$$

$$= I\dot{\omega} = \frac{Mr^2}{2}\dot{\omega}$$

$$= \frac{16}{2}\frac{(0.5)^2}{4} \times \frac{\pi}{2} = \frac{\pi}{4} \text{ Nm.}$$

Rate at which work is done
= (Torque) × (Angular velocity)

$$= \frac{\pi}{4} \times 4\pi$$

$$= \pi^2 \text{ watts.}$$

3.23 Analogy Between Linear Motion and Rotational Motion

We have noticed at every stage the analogy between linear motion and rotational motion. Table 3.1 summarizes the analogy showing the quantities relating to linear motion in one column and the corresponding quantities relating to rotational motion in another column.

Table 3.1

Linear Motion	Rotational Motion
Displacement s	Angular displacement θ
Velocity v	Angular velocity ω
Acceleration a	Angular acceleration $\dot{\omega}$
Mass m	Moment of inertia I ($I = \Sigma mr^2$)
Force F	Torque τ
$F = m.a$	$\tau = I\dot{\omega}$
Kinetic energy $= \frac{1}{2}mv^2$	Kinetic energy $= \frac{1}{2}I\omega^2$
Work done $= Fs$	Work done $= \tau\theta$
Momentum $P = mv$	Angular momentum $L = I\omega$

EXERCISES

3.1 Why does a cyclist lean to one side while going along a curve? In which direction does he lean?

3.2 A car is taking a sudden turn to the left. A passenger in the front seat finds himself sliding towards the door. Explain, indicating the forces acting on the passenger and on the car at this instant.

3.3 According to Newton's universal law of gravitation, every body in this universe attracts every other body. But we do not see bodies on the surface of the earth moving towards one another on account of this force of attraction. Why?

3.4 Imagine a space-craft going from the earth to the moon. How does its weight vary as it goes from the earth to the moon? Will there be any change in its mass?

3.5 What is gravitational field strength? What are its units?

3.6 What are the conditions under which a rocket fired from the earth becomes a satellite of the earth and orbits in a circle?

3.7 If an earth satellite is put in orbit at a height where the resistance due to the atmosphere cannot be neglected, how will the motion of the satellite be affected?

3.8 The melting of the polar icecaps is given as a possible cause of the variation in the period of rotation of the earth about its axis. Explain.

3.9 A body weighing 0.4 kg is whirled in a vertical circle making 2 revolutions per second. If the radius of the circle is 1.2 m, find the tension in the string when the body is (a) at the top of the circle; (b) at the bottom of the circle.

(71.86 N; 79.70 N)

3.10 A string can withstand a tension of 25 N. What is the greatest speed at which a body of mass 0.5 kg can be whirled in a circle using 0.5 m length of the string. Neglect the force of gravity on the body.

($5\ ms^{-1}$)

3.11 An aeroplane travelling at a speed of 500 km h^{-1} banks at an angle of 30° as it makes a turn. What is the radius of the curve?

(3.41×10^3 m)

3.12 A train has to negotiate a curve of 400 m. By how much should the outer rail be raised with respect to the inner rail for a speed of 48 km h^{-1}? The distance between the rails is 1 metre.

(0.0454 m)

3.13 Compare the gravitional acceleration of the earth due to attraction of the sun with that due to attraction of the moon.

(1790 : 1)

3.14 The radius of the moon is 1.7×10^6 m and its mass is 7.35×10^{22} kg. What is the acceleration due to gravity on the surface of the moon?

($1.70\ ms^{-2}$)

3.15 The mass of the planet Jupiter is 1.9×10^{27} kg and that of the sun 1.99×10^{30} kg. The

mean distance of the sun from Jupiter is 7.8×10^{11}. Calculate the gravitational force which the sun exerts on Jupiter. Assuming that Jupiter moves in a circular orbit around the sun, calculate the speed of Jupiter.

$(4.146 \times 10^{23}\ \text{N}; 1.3 \times 10^{4}\ \text{ms}^{-1})$

3.16 Calculate the escape velocity on the surface of the moon. Given, the radius of the moon $= 11.7 \times 10^{6}$ m; the mass of the moon $= 7.35 \times 10^{22}$ kg.

$(2.4 \times 10^{3}\ \text{ms}^{-1})$

3.17 What is the angular velocity in rad/s of a flywheel making 300 rpm?

$(10\pi\ \text{rad/s})$

3.18 A thin metal hoop of radius 0.25 m and mass 2 kg starts from rest and rolls down an inclined plane. If its linear velocity on reaching the foot of the plane is 2 ms^{-1}, what is its rotational kinetic energy at that instant?

(4 J)

3.19 A solid cylinder rolls down an inclined plane. Its mass is 2 kg and radius 0.1 m. If the height of the inclined plane is 4 m, what is its rotational kinetic energy when it reaches the foot of the plane? Assume that the surfaces are smooth.

(26.13 J)

3.20 On the application of a constant torque, a wheel is turned from rest through an angle of 200 radians in 8 s. (a) What is its angular acceleration? (b) If the same torque continues to act, what will be the angular velocity of the wheel after 16 s from start?

$(6.25\ \text{rad.}\ \text{s}^{-2}; 100\ \text{rad.}\ \text{s}^{-1})$

3.21 A cord is wound around the circumference of a wheel of diameter 0.3 m. The axis of the wheel is horizontal. 0.5 kg mass is attached at the end of the cord and it is allowed to fall from rest. If the weight falls 1.5 metres in 4 s, what is the angular acceleration of the wheel? Also, find out the moment of inertia of the wheel.

$(1.25\ \text{rad.}\ \text{s}^{-2}; 0.588\ \text{kg m}^{2})$

3.22 A flywheel of mass 25 kg has a radius of 0.2 m. It is making 240 rpm. What is the torque necessary to bring it to rest in 20 s? If the torque is due to a force applied tangentially on the rim of the flywheel, what is the magnitude of the force?

$(-\pi/5\ \text{Nm}; \pi\ \text{N})$

UNIT 4

Harmonic Oscillations

If a body moves back and forth repeatedly about a mean position, its motion is called an *oscillation*. As examples, we may think of the motion of the pendulum of a wall-clock, the oscillation of a mass suspended on a spring, the vibration of the drum of a tabla or the wire of a sitar, the angular swinging of a body suspended on a wire, etc. It may be remarked that the motion of the planets around the sun is periodic motion, but not oscillatory, because the idea of 'back and forth' is absent there. The motion of the balance-wheel of a watch is an oscillation. An oscillation is (by and large) periodic motion, but not every type of periodic motion is oscillation.

An essential concept in oscillations is that of a position or state of equilibrium. If a system is disturbed from its position of equilibrium and left to itself, it oscillates about the mean position. All oscillations can be expressed in terms of harmonic functions (e.g., sine and cosine functions. An oscillation which can be expressed in terms of one single harmonic function is called a *harmonic oscillation* and in this unit we shall discuss various aspects of harmonic oscillations. All other oscillations can be easily understood in terms of harmonic oscillations. For this reason, its discussion is of importance in mechanics, where disturbances from a state of equilibrium are to be studied; also in sound and light, where harmonic waves are generated by oscillators; in electrical circuits, where oscillating currents of varying frequencies are produced and then electromagnetic waves (e.g., radio waves) are generated in space.

4.1 The Description of a Harmonic Oscillation

We shall express the displacement of a body from its position of equilibrium by the symbol ξ, and describe an oscillation through a (ξ, t) graph. In Fig. 4.1, three such graphs are shown. Each one shows ξ varying repeatedly between some limits and, therefore, represents an oscillation. For Fig. 4.1b, we can write a simple relation for ξ, t

$$\xi = a \cos 2\pi \frac{t}{T}. \qquad \ldots (4.1)$$

For Fig. 4.1c also, we can write a simple relation as:

$$\xi = a \sin 2\pi \frac{t}{T}. \qquad \ldots (4.2)$$

Both these curves are said to represent a *harmonic oscillation*.

In fact, the equation (4.2) can be written as:

$$\xi = a \cos \left(2\pi \frac{t}{T} - \frac{\pi}{2} \right). \qquad \ldots (4.3)$$

Or, in the graph we may shift the curve (c) by T/4 backward on the time-scale to make it fit the curve (b). Therefore, we may represent a harmonic oscillation in general by

$$\xi = a \cos \left(2\pi \frac{t}{T} + \phi_0 \right). \qquad \ldots (4.4)$$

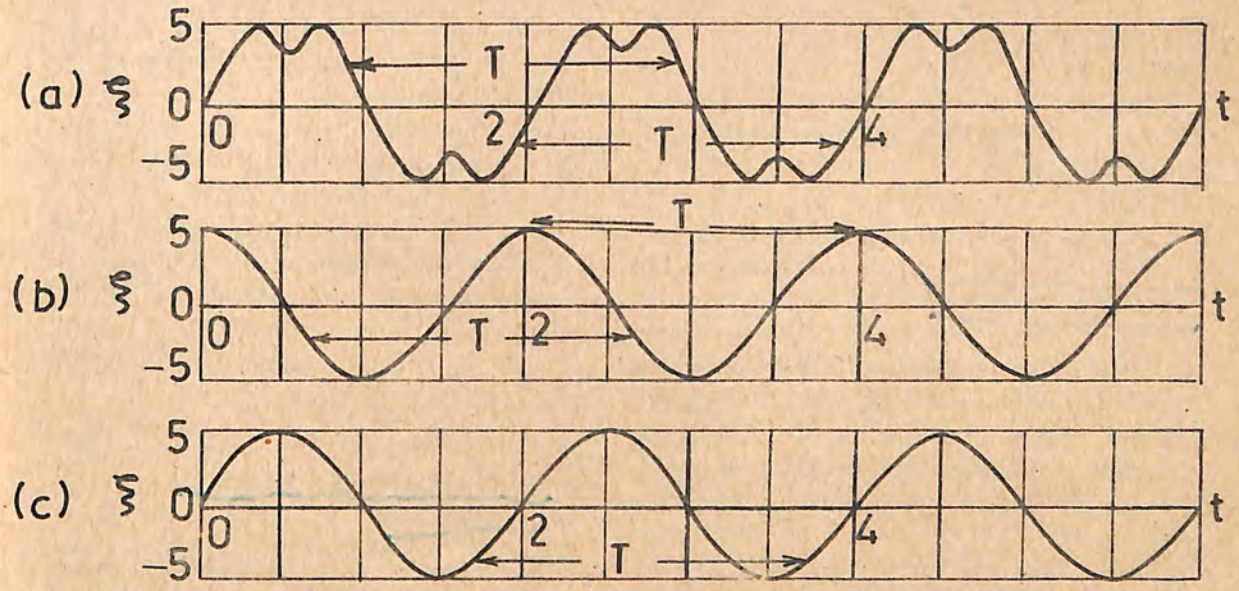


Fig. 4.1 Periodic oscillations: (a) not harmonic; (b) harmonic; (c) the same as in (b), but the phase $\pi/2$ behind it.

We shall now define some quantities characterizing a harmonic oscillation.

Amplitude

The quantity a in the equation (4.4) represents the maximum magnitude of the displacement ξ; actually, ξ varies between $+a$ and $-a$. This quantity a is called the *amplitude* of the harmonic oscillation.

In general, the quantity ξ could be a linear displacement, an angular displacement (as in the balance-wheel of a watch), an electrical charge (as in an electrical circuit), a temperature (as in thermal oscillations). We can look upon the equation (4.4) in a more general sense, and hence think of a as the amplitude of the corresponding quantity described by ξ. In linear mechanical oscillations, the unit of ξ (or a) is metre and in angular oscillations, it is radian.

Period

The smallest interval of time at which the details of motion repeat is called the *periodic time* or *period*, represented by T. In Fig. 4.1, the periodic time is shown for all the three curves. In fact, for a given curve one may start from any point and search for the nearest point (on the time-scale) where the curve starts repeating; the interval gives T.

In the equation (4.4), we may see that changing t to $t + T$ makes ξ identical. Formally, we may write

$$\xi(t + T) = \xi(t) \qquad \ldots (4.5)$$

as a way of defining the period. The dimension of a period is generally time*.

Frequency

The number of oscillations completed per unit time is called *frequency*, denoted by ν. Obviously,

$$\nu = \frac{1}{T} \qquad \ldots (4.6)$$

* We could use the equation (4.4) to describe the harmonic variation of ξ with respect to any parameter—like the variation of vertical height with distance in the wave patterns on sand in a desert. In that case, t will represent horizontal distance.

and the unit of ν would be (second)$^{-1}$, written as s^{-1}. However, quite often the expression 'cycles per second' (cps) is also used. The name Hertz (Hz) has been given to s^{-1}.

Thus, $500\ s^{-1} = 500\ Hz = 500$ cps.

Angular Frequency

The quantity $2\pi/T$, which equals $2\pi\nu$, is often used for convenience and is given the name *angular frequency*. It is simply the frequency ν multiplied by the numerical quantity 2π. Angular frequency is denoted by ω, so that*

$$\omega = 2\pi/T = 2\pi\nu \qquad \ldots (4.7)$$

One can now rewrite the equation (4.4) as:

$$\xi = a\cos(2\pi\nu t + \phi_o)$$
$$= a\cos(\omega t + \phi_o) \qquad \ldots (4.8)$$

Phase

The argument of the cosine in the equation (4.4) is called the *phase* of the oscillation at time t, and is denoted by ϕ. Thus,

$$\phi = 2\pi \frac{t}{T} + \phi_o \qquad \ldots (4.9)$$

The phase continuously increases with time, and one may write it variously, as

$$\phi = \frac{2\pi}{T}\Delta t$$
or $$= 2\pi\nu.\Delta t$$
or $$= \omega\Delta t. \qquad \ldots (4.10)$$

One may use the equation (4.9) to define the period T as that time in which the phase changes by 2π.

The quantity ϕ_o is called the *initial phase*, or the phase at time $t = 0$. We have seen earlier that if we describe a harmonic wave using a sine function, the phase differs by $\pi/2$. For this reason, it is a useful practice to use always one form, and we shall use the cosine form** The initial phase really depends upon where we choose to fix the zero for time measurement.

4.2 The Dynamics of a Harmonic Oscillation

If a body is in equilibrium, the forces on it balance. If it is now displaced from its state of equilibrium by an amount ξ, there are three possibilities: (a) the forces may still be in balance, (b) the net force produced may be in the same direction as ξ, and (c) the net force produced may be in a direction *opposite* to that of ξ. The equilibrium is called 'neutral' in case (a), 'unstable' in case (b) and 'stable' in case (c). Our interest is in stable equilibrium and the disturbances created in it. The net force produced by the displacement ξ is always opposite to the direction of ξ and hence is called a *restoring force*. Whether we depress the branch of a tree, or bend a metre-scale, or stretch a spring, or press a rubber ball, or move aside the bob of a pendulum—in all these cases, a restoring force is produced. The magnitude of the force is, in general, some function of ξ. But whatever the function, the result is that under this force the body returns towards its position of equilibrium, overshoots because of the kinetic energy gained by the time it reaches $\xi = 0$, then loses the speed because of the restoring force till the kinetic energy becomes zero,

* It may be remarked that in circular motion (like that of a planet), a term *angular velocity* occurs, represented by the same symbol ω as used for angular frequency. Now, angular velocity has the dimension of angle/time. If we take angle as dimensionless, then the dimension becomes s^{-1} which is the same as that for angular frequency. The relation of angular velocity to the number of circles completed is also the same as in the equation (4.7).

** The wave-forms 4.1b as well as 4.1c are called *sinusoidal*.

and so on. It oscillates back-and-forth about the mean position.

The period of oscillation for a given system may, in general, depend on the amplitude of the oscillation. But in the case of our interest here we shall find that the period is independent of the amplitude. If ξ is small enough, the magnitude of the restoring force is proportional to ξ. For this case, we write:

$$F = -K\xi \qquad \ldots (4.11)$$

where the minus $(-)$ sign shows that the directions of F and ξ are always opposite, and K is a constant of the system, called *force constant* and having the unit N/m.

If m is the mass of a body, then by Newton's law we write for the acceleration:

$$\ddot{\xi} = \frac{\text{Force}}{\text{Mass}} = -\frac{K}{m}\xi. \qquad \ldots (4.12)$$

We will find it convenient to use a substitution

$$\omega^2 = \frac{K}{m} \qquad \ldots (4.13)$$

and write the equation (4.12) as:

$$\ddot{\xi} = -\omega^2 \xi. \qquad \ldots (4.14)$$

The solution of this equation is somewhat difficult. But we can see that the result is of the form

$$\xi = a\omega \cos(\omega t + \phi_0). \qquad \ldots (4.15)$$

Differentiating the equation (4.15) two times gives

$$\dot{\xi} = -a \sin(\omega t + \phi_0)$$

$$\ddot{\xi} = -a\omega^2 \cos(\omega t + \phi_0)$$

or $$\ddot{\xi} = -\omega^2 \xi$$

which is the same as the equation (4.14). Thus, the equation (4.15) is a solution of the equation (4.12) with $\omega = \sqrt{K/m}$.

The motion described by the equation (4.15) has already been named as a harmonic oscillation. Now, we see the dynamic condition which is responsible for producing motion of that kind. Harmonic oscillation is that in which the force acting on a system is always against its displacement ξ from a mean position and proportional to ξ in magnitude. The equation (4.11) is the mathematical form of this definition. One could also use the equation (4.14) and define harmonic oscillation as that in which the acceleration is always proportional to the displacement and against it in direction. In this case, the square-root of the constant of proportionality gives the angular frequency.

We now see that

$$\nu = \omega/2\pi = \frac{1}{2\pi}\sqrt{\frac{K}{m}} \qquad \ldots (4.16)$$

$$T = 1/\nu = 2\pi\sqrt{m/K}. \qquad \ldots (4.16a)$$

Harmonic oscillations could also be angular. In that case, a restoring couple C would be produced, and, in Newton's law, we shall need moment of inertia I, instead of mass. Thus, the equation (4.16) will take the form

$$\nu = \frac{1}{2\pi}\sqrt{\frac{C}{I}} \qquad \ldots (4.17)$$

Thus, depending on what kind of oscillation we discuss, the quantities corresponding to K and m of the equation (4.16) will go on taking different names and forms. As a general nomenclature, K is called the *spring factor* and m the *inertia factor*, and the equation (4.16) then becomes

$$\nu = \frac{1}{2\pi}\sqrt{\frac{\text{Spring factor}}{\text{Inertia factor}}} \qquad \ldots (4.18)$$

4.3 Some Examples of Harmonic Oscillations

We shall now examine some typical cases in which harmonic oscillations occur. In each case, some simplifying assumptions are made, which should be noticed.

Mass Oscillating on a Spring

The elastic force produced in a spring when it is stretched or compressed is of the character $F = -K\xi$, if the total size-change is (say) within 10 per cent of the length of the spring. However, some special springs have been produced in which this relation closely holds for even 100 per cent stretching.

Fig. 4.2 shows at (a) the spring alone, at (b) the body P attached and in a state of equilibrium, and at (c) the body displaced to the right by an amount ξ relative to the state (b). We may imagine the body as lying on a frictionless horizontal table, so that gravitational pull on it is balanced by the table's reaction force and the only force is the elastic force of the spring.

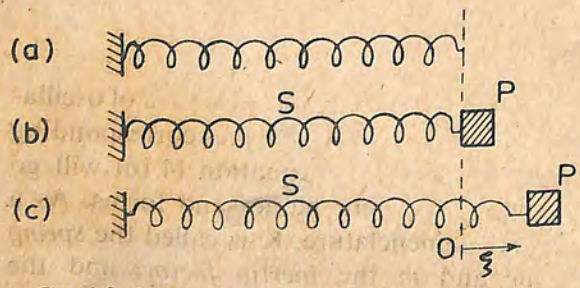


Fig. 4.2 Oscillations of a body P on a spring S.

If the mass of the spring is negligible compared with the mass m of the body P, and if the left end of the spring is taken as attached to a body of the mass much greater than m, the oscillation frequency is given simply by

$$\nu = \frac{1}{2\pi}\sqrt{\frac{K}{m}} \qquad \ldots (4.19)$$

If the oscillations are vertical – in which case we could use a rubber band also in place of a spring – the result is the same, except that the equilibrium state is not the unstretched spring, but a spring already stretched by the amount d (Fig. 4.3). The frequency is still the same and not depending on the acceleration due to gravity at all.

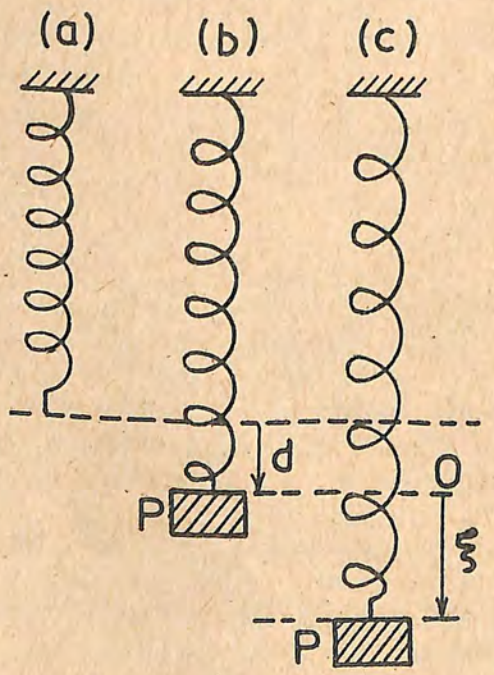


Fig. 4.3 Vertical oscillations of a body on a spring.

Simple Pendulum

We may tie a piece of stone with a thread and hang it on some peg to get a pendulum. The pendulum of a wall-clock is not like this because its whole length is

rigid (with the lower end heavier, of course); on the other hand, the 'simple pendulum' is also not like this because the 'thread' is taken as an ideal massless and inelastic one, and the bob' (the stone) is taken as a point-mass. The support is taken to be rigid and of infinite mass.

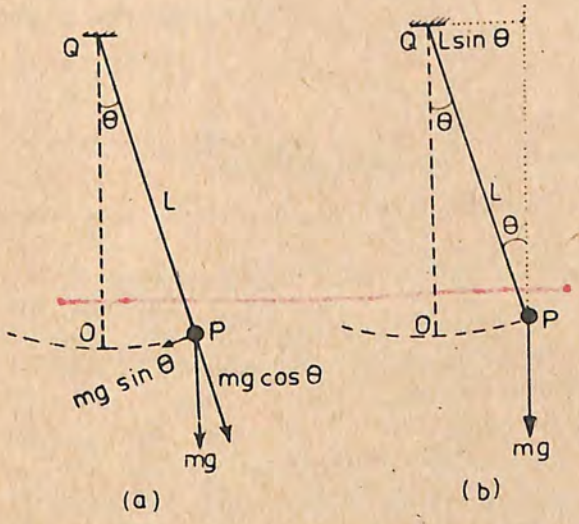


Fig. 4.4 A simple pendulum examined in two ways: (a) as the linear motion of the bob P; (b) as the angular motion of the thread and bob Q P about Q.

The simple pendulum (Fig. 4.4) can be examined in two ways. In Fig. 4.4a, the gravitational pull mg, which is vertically downwards, is shown in two components: mg cosθ along the thread is balanced by the tension T in the thread, and mg sinθ perpendicular to the thread is the only net force acting on the mass. The motion of the bob is along an arc of a circle, but for small amplitude one takes the path to be straight and mg sinθ to be equal to mgθ. Thus, force = mgθ, displacement = Lθ, spring factor = mg/L. The mass factor is m, the mass of the bob. Hence,

$$\nu = \frac{1}{2\pi}\sqrt{\frac{mg/L}{m}} = \frac{1}{2\pi}\sqrt{\frac{g}{L}} \qquad \ldots(4.20)$$

A notable fact in this case is that m appears in both the spring factor and the mass factor and, hence, cancels out*.

In Fig. 4.4b, we take the moment of mg about the support point Q that gives a torque mg L sinθ which becomes mg Lθ for small θ. The moment of inertia of the system QP about Q is mL^2, since the string has no mass. We get

$$\nu = \frac{1}{2\pi}\sqrt{\frac{\text{Spring factor mgL}}{\text{Inertia factor mL}^2}}$$

$$= \frac{1}{2\pi}\sqrt{\frac{g}{L}} \qquad \ldots(4.21)$$

which is the same as the equation (4.20). This treatment is advantageous in the sense that the approximation is at just one place (taking $\sin\theta = \theta$). For oscillations of large amplitude, one can, therefore, search for an improved solution. Also, for rigid pendulums this can be adapted.

Balance-wheel of a Watch

In a watch, a balance-wheel is held between two 'diamond points', a hair-spring controlling its angular oscillations about a mean equilibrium position. For an angular displacement θ, the hair-spring develops a torque, Cθ, and if the moment of inertia of the balance-wheel about the axis of oscillation is I, we get

$$\nu = \frac{1}{2\pi}\sqrt{\frac{C}{I}} \qquad \ldots(4.22)$$

Liquid in a U-tube

Fig. 4.6 shows a liquid in a U-tube disturbed from the state of equilibrium by an amount ξ. Assuming equal cross-sections A on both sides, the force produced is $-2\xi A\rho g$,

* Actually, m in mg/L is the gravitational mass, while m in the denominator is the inertial mass.

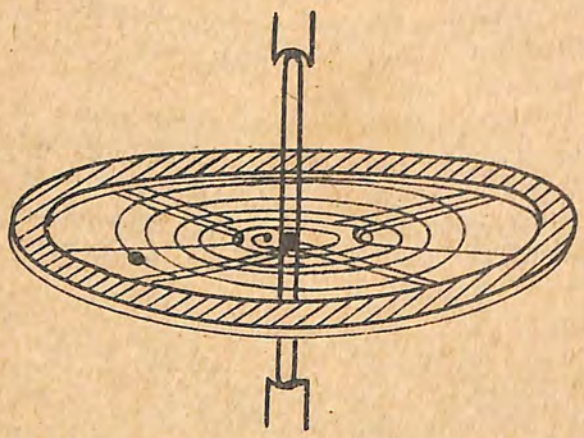


Fig. 4.5 The balance-wheel and the hair-spring of a watch.

where ρ is the density of the liquid and g the acceleration due to gravity, and the minus sign shows that if the right arm liquid rises by ξ, the force would try to depress it. The force has the character $-K\xi$, with $K = 2A\rho g$. Hence, the liquid will oscillate harmonically with angular frequency $\omega = \sqrt{2A\rho g/A\rho L}$, where $A\rho L$ is the total mass of the liquid in length L, assuming uniform cross-section. The result is $\omega = \sqrt{2g/L}$, and one notices that ω does not depend on the cross-section of the tube or the density of the liquid.

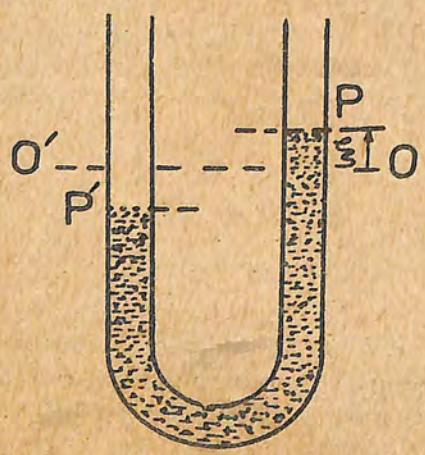


Fig. 4.6 A liquid oscillating in a U-tube; 'OO' is the equilibrium state; PP' is the displaced state.

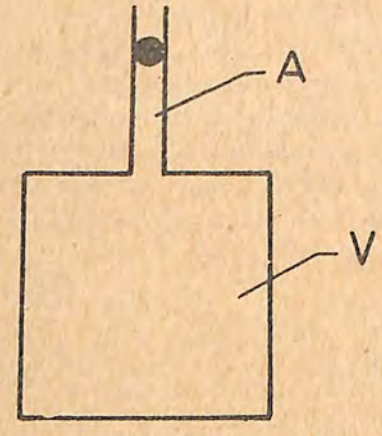


Fig. 4.7 A ball in the neck of an air chamber.

The Case of an Air Chamber with a Neck

Fig. 4.7 shows an air chamber of volume V, having a neck of the area of cross-section A and a ball of mass m fitting smoothly in the neck. If the ball is pressed down by the distance ξ, the air volume decreases by ξA and an excess pressure p is produced, as given by

$$p = E\frac{\xi A}{V}$$

where E is the elasticity of air. The force produced is upward (opposite to ξ) and is pA. So, we get $F = -(EA^2/V)\,\xi$. This has the character $-K\xi$; hence, the ball executes harmonic oscillations. The angular frequency ω is $\sqrt{EA^2/Vm}$, assuming that the mass of air in the neck is negligible compared with that of the ball.

But there is the other extreme case, when there is no ball at all and only the air in the neck oscillates under the 'spring' action of the air in the chamber. In that case if the neck has the length L and the air density is ρ, we have $m = AL\rho$ and, hence, $\omega = \sqrt{EA/VL\rho}$.

EXAMPLE 4.1

A spring compressed by 0.1 m develops a restoring force 10N. A body of mass 4 kg is placed on it. Deduce (i) the force constant of the spring, (ii) the depression of the spring under the weight of the body (take g = 10N/kg), and (iii) the period of the oscillation, if the body is disturbed.

Solution

$$K = \frac{\text{force}}{\text{displacement}} = \frac{10\ \text{N}}{0.1\ \text{m}}$$

$$= 100\ \text{N/m}.$$

$$\text{displacement} = \frac{\text{force}}{K}$$

$$= \frac{4\ \text{kg} \times 10\ \text{N/kg}}{100\ \text{N/m}}$$

$$= 0.4\ \text{m}.$$

$$T = 2\pi\sqrt{\frac{m}{K}} = 2\pi\sqrt{\frac{4\ \text{kg}}{100\ \text{N/m}}} = \frac{2\pi}{5}\ \text{s}.$$

EXAMPLE 4.2

In a HCl molecule, we may treat Cl to be of infinite mass and H alone oscillating. If the oscillation of the HCl molecule shows frequency $9 \times 10^{13}\ \text{s}^{-1}$, deduce the force constant. The Avogadro number $= 6 \times 10^{26}$ per kg-mole.

Solution

$$\nu = \frac{1}{2\pi}\sqrt{\frac{K}{m}}$$

$$K = 4\pi^2 \nu^2 m$$

$$= 4(3.14)^2 (9 \times 10^{13})^2 \cdot \frac{1}{6 \times 10^{26}}$$

$$= 5.4 \times 10^2\ \text{N/m}.$$

EXAMPLE 4.3

A pendulum clock shows accurate time. If the length increases by 0.1 per cent, deduce the error in time per day.

Solution

The correct number of seconds per day is 86400. If the incorrect number given by the clock is 86400 + x, we see from the equation (4.17) that

$$\frac{86400 + x}{86400} = \sqrt{\frac{L}{L + .001L}}$$

$$= (1 + .001)^{-\frac{1}{2}}$$

$$= 1 - \frac{1}{2}(.001)$$

where L is the true length and 0.001L is the increase in it.

$$\frac{x}{86400} = -\frac{1}{2}(.001)$$

$$x = -43\ \text{s}.$$

The clock goes slower by 43 seconds each day.

4.4 Particle Velocity and Acceleration

For the harmonic oscillation represented by the displacement equation

$$\xi = a\cos(\omega t + \phi). \qquad \ldots(4.23)$$

the velocity equation is

$$\dot{\xi} = -\omega a \sin(\omega t + \phi)$$

$$= a\omega\cos(\omega t + \phi + \pi/2) \qquad \ldots(4.24)$$

and the acceleration equation is

$$\ddot{\xi} = -\omega^2 a\cos(\omega t + \phi)$$

$$= \omega^2 a\cos(\omega t + \phi + \pi). \qquad \ldots(4.25)$$

Comparing these equations, the following facts stand out:

(i) The velocity amplitude is ω times a; the acceleration amplitude is ω^2

times a.

(ii) The velocity is $\pi/2$ ahead of the phase relative to the displacement, the acceleration is further $\pi/2$ ahead of the phase.

'Ahead of phase by $\pi/2$' is equivalent to saying 'ahead of time by T/4', since one period corresponds to phase 2π. Therefore, in graphical representation the velocity-time graph is shifted T/4 towards the earlier time, and acceleration-time graph is shifted by further T/4 towards the earlier time relative to the ξ, t graph. Fig. 4.9 shows the three graphs together, using one time axis but different scales on the y-axis for ξ, $\dot{\xi}$ and $\ddot{\xi}$.

If we use v_o for the velocity amplitude and g_o for the acceleration amplitude, the result (i) amounts to

$$v_o = \omega a; \quad g_o = \omega^2 a. \qquad \ldots (4\,26)$$

Using the frequency ν the same relations become

$$v_o = 2\pi\nu a; \quad g_o = 4\pi^2\nu^2 a. \qquad \ldots (4.27)$$

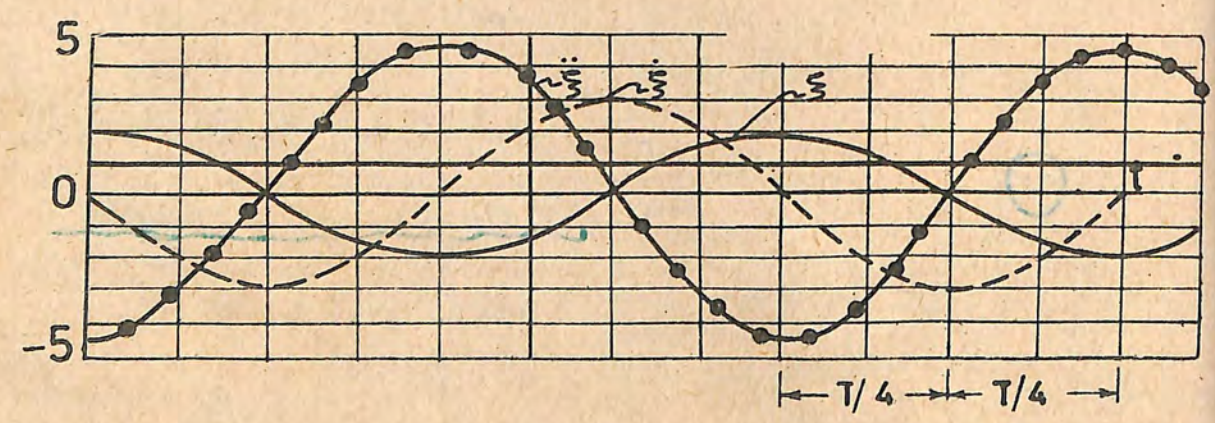


Fig. 4.9 ξ, $\dot{\xi}$ the ξ for the same vibration plotted against the time t. If ω = 1500 Hz, and the scale for ξ is 1 unit = 10^{-2} cm, that for $\dot{\xi}$ will be 1 unit = 10 cm/sec, and that for $\ddot{\xi}$ will be 1 unit = 10^4 cm/sec^2.

EXAMPLE 4.4

In a pendulum oscillation, the amplitude is 0.05 m and the period 2.0 s. Compute the maximum velocity.

Solution

$a = 0.05$ m $\quad \nu = 1/T = 0.5\ s^{-1}$.

$v_0 = 2\pi \times 0{\cdot}5 \times {\cdot}05 m\ s^{-1}$

$= 0.16\ m\ s^{-1}$

EXAMPLE 4.5

In a ballistic pendulum, a gunshot hits the bob of a pendulum horizontally and the amplitude of the pendulum gives the velocity of the bullet. In one such case, a 2.0 g bullet fired got struck into a 500 g sand bag, used as the bob, and the amplitude of oscillation became 20 cm. The pendulum makes 20 oscillations per minute. Deduce the bullet velocity.

Solution

$$a = 20\text{ cm}, \qquad \nu = \frac{20}{60} = \frac{1}{3}\text{ s}^{-1}$$

$$v_o = 2\pi \times \frac{1}{3} \times 20 = 40\pi/3\text{ cm s}^{-1}$$

If the bullet velocity was v, by conservation of linear momentum,

$$2.0\,v = (2 + 500)\,v_o$$

$$v = \frac{502}{2.0} \times \frac{40\pi}{3}\text{ cm s}^{-1}$$

$$= 105\text{ m s}^{-1}.$$

4.5 Energy in Harmonic Oscillations

In a harmonic oscillator, the energy is partly kinetic E_k and partly potential E_p. When the displacement is zero, the potential energy also is zero, because the equilibrium is taken as the state relative to which the potential energy is measured. When the displacement is extreme ($+a$ or $-a$), the kinetic energy is zero, because the velocity becomes zero. All energy must, therefore, become potential. By the law of conservation of energy, the total energy E must remain constant, because we are presently ignoring frictional losses. Therefore, the total energy of oscillation E equals the maximum value of kinetic energy, which we denote by E_k^o. Thus,

$$E = E_k^o$$

$$\text{or} \quad E = \frac{1}{2}mv_o^2 = 2\pi^2 \nu^2 a^2 m \quad \ldots(4.28)$$

where v_o is the velocity amplitude, and we have used the equation (4.27) The maximum of potential energy E_p^o must also be $2\pi^2\nu^2a^2m$.

If the oscillations are angular, then m is replaced by the moment of inertia of the oscillator about the axis of rotation, and v_o is replaced by the amplitude of angular velocity θ_o. Thus,

$$E = 2\pi^2 \nu^2 \theta_0^2 I. \quad \ldots(4.29)$$

For the general value of potential energy, we have a simple result:

$$E_p = E - E_k = E_k^o - E_k$$

$$= \frac{1}{2}m(v_o^2 - v^2), \quad \ldots(4.30)$$

We leave it to the reader to use the equations (4.23), (4.24) and $\omega^2 = K/m$ to show that

$$E_p = \frac{1}{2}K\xi^2;$$

$$E = E_p^o = \frac{1}{2}Ka^2. \quad \ldots(4.31)$$

This is in agreement with the equation (4.28) because K is equal to $\omega^2 m$, which is $4\pi^2\nu^2 m$.

EXAMPLE 4.6

Two bodies A and B (Fig. 4.10) of mass 1 kg and 2 kg, respectively are rigidly connected to a spring of force constant 400 N/m.

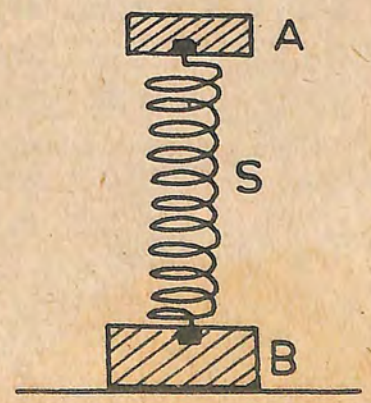


Fig. 4.10

The body B rests on a horizontal table. From the rest position, the body A is compressed by a distance d = 2 cm and then released. Deduce (i) the frequency of oscillation, (ii) the maximum velocity of the body A, (iii) the amplitude of the harmonic variation of reaction of the table on the body B.

Solution

$$\text{Frequency } \nu = \frac{1}{2\pi}\sqrt{\frac{K}{m}} = \frac{1}{2\pi}\sqrt{\frac{400}{1}}$$

$$= \frac{10}{\pi} \text{ Hz.}$$

$$\text{Oscillation energy } E = \frac{1}{2} Kd^2$$

$$= 200 \times \left(\frac{2}{100}\right)^2 = \frac{2}{25} \text{ J.}$$

Mean reaction R_o = 3 kg wt = 30 N.

As *A* oscillates, the tension in the spring S does not remain equal to 1 kg wt = 10 N all through. For compression of 2 cm, extra tension = Kx = 8 N. At the other extreme, there will be rarefaction of 2 cm, and tension lower by 8 N.

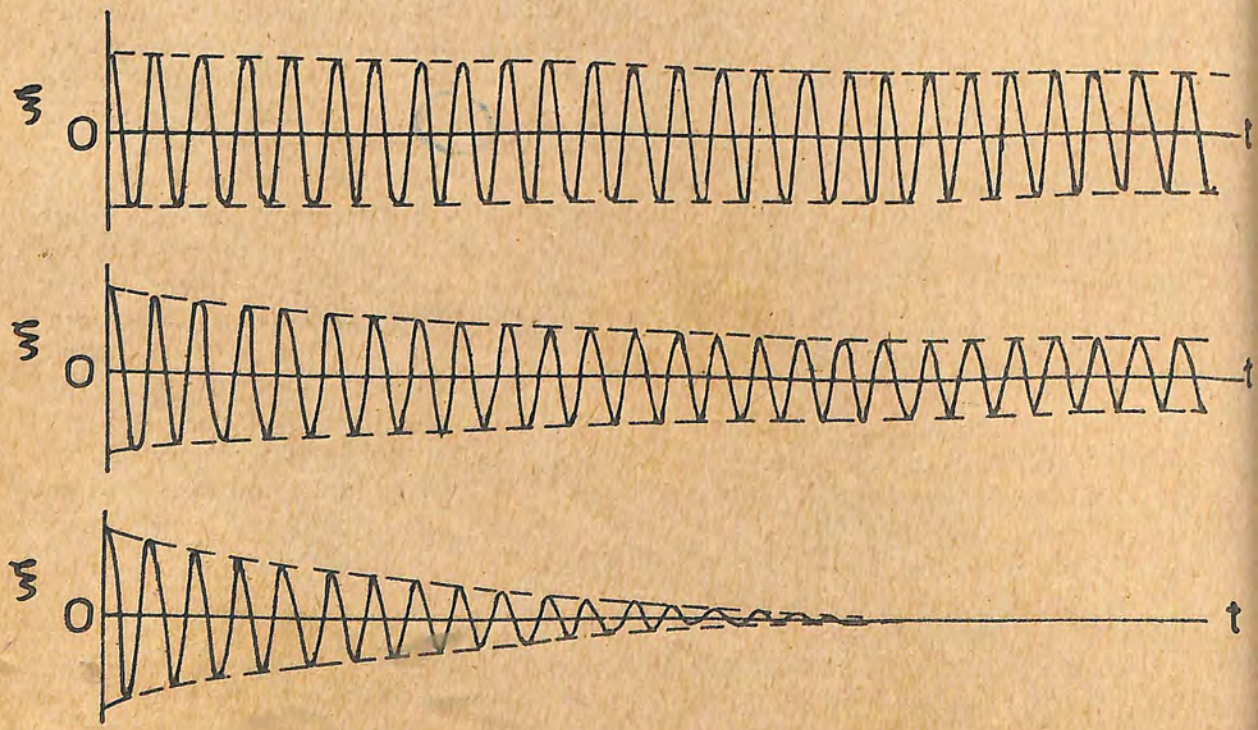
∴ amplitude of variation of R = 8 N

i.e., reaction varies between (30 + 8) N and (30 − 8) N.

4.6 Resonance

We may initiate an osillation by imparting to the system some initial displacement ξ, or some initial velocity v_o. In the first case, the energy is initiated in the potential form $E = \frac{1}{2} K \xi^2$; in the second case, it is initially fed in the kinetic form $E = \frac{1}{2} mv_o^2$. Left to itself thereafter, the system in either case oscillates with a frequency of its own. It is

Fig. 4.11 Damped oscillations of small, medium and large dampings.



called a *free oscillation* in the sense that the system is left free after feeding some energy initially. For its frequency, we write

$$\nu_o = \frac{1}{2\pi}\sqrt{\frac{K}{m}}$$

the subscript o being added to express the idea of free oscillations.

In practice, the system always will have some resistance, i.e., a cause for the loss of energy to the surrounding or change into heat form within the 'spring' itself. As a result, the energy of oscillations, and hence the amplitude, falls as time passes (Fig. 4.11). These are called *damped oscillations*. But we can feed energy to the oscillator to keep it oscillating. If this energy is fed continuously, the oscillation frequency is still close to ν_o—all that occurs is that the rate of feeding energy just balances the rate of loss of oscillation energy, i.e., the power fed equals the power lost by friction. These are called *maintained oscillations*. An equation like

$$\xi = a\cos(\omega t + \phi_o),$$

with the amplitude constant, really represents this case or an idealised case of free oscillation.

The energy from an external agent can also be fed with a definite frequency ν. We shall call the external agent the driver, so that ν is the frequency of the driver. The system under consideration is then called the driven, with its natural frequency ν_o. In such a case, with the passage of time, the natural (free) vibration of frequency ν_o dies out and the system starts oscillating with the frequency ν of the driver. Oscillations of the frequency ν gain in amplitude with time till a stage is reached where the power lost equals the power fed. These are called *forced oscillations*: their frequency is ν not ν_o.

The power fed in forced oscillations is very small if ν and ν_o have a large difference. But it becomes quite large as ν approaches ν_o; in that situation, the amplitude of forced oscillation becomes quite large. The case $\nu \approx \nu_o$ is called *resonance*. We say that the driver and the driven are in resonance, or that the two systems are in resonance. If there are drivers of many frequencies, the driven system accepts energy most favourably from just that driver which matches it in frequency. This is what happens when we tune our radio to a particular 'station'. Waves from all the stations are present around the antenna, but the radio 'selects' just that station for which we set the frequency of the radio circuit.

On occasions, resonance can be disastrous. When an army marches on a bridge, it is possible that the frequency of footsteps may match one of the natural frequencies of the bridge and then the bridge may pick up a large amplitude of oscillation. So while crossing the bridge, the army is not made to march in rhythm. On a wavy road sometimes when a car-body oscillations (on its springs) match in frequency the jerks from the wavy road, the driver at once alters the speed of the car. Within the earth some periodic oscillations are often present—too small to be significant. But if the frequency of these happens per chance to match, for instance, the frequency of natural waves on a lake, or the oscillations of a structure, they may gain large amplitudes, causing disaster.

EXERCISES

4.1 The bodies of mass 1, 4, 9, 16 kg are oscillating in turn on a spring of force constant 100 N/kg. Deduce the angular frequencies.

10, 5, 10/3, 5/2 Hz)

4.2 A body of mass 1 kg is made to oscillate in turns on springs of force constants 1, 4, 9, 16 N/kg. Deduce the angular frequencies. (1, 2, 3, 4 Hz)

4.3 The bottom of a dip on a road has a radius of curvature R. A rickshaw of mass M left a little away from the bottom oscillates about this dip. Deduce an expression for the period of oscillation. ($T = 2\pi\sqrt{R/g}$)

4.4 On the same (ξ, t) graph, show two-harmonic oscillations A and B such that amplitudes are in the ratio 1 : 2 and frequencies in the ratio 1 : 3. Deduce the curve arising from the addition of ξ_A and ξ_B.

4.5 A harmonic oscillation is represented by $\xi = 0.34 \cos (3000\ t + 0.74)$ where ξ and t are in mm and s, respectively. Deduce (i) the amplitude, (ii) the frequency and angular frequency, (iii) the period, and (iv) the initial phase.

(0·34 m.m.; 1500/π Hz. 3000 Hz: π/1500 s; 0.74 rad)

4.6 Two identical springs, each of a spring factor K, may be connected in the following different ways. Deduce the spring factor for the oscillation of the body P in each case. (2 K; K/2; 2 K)

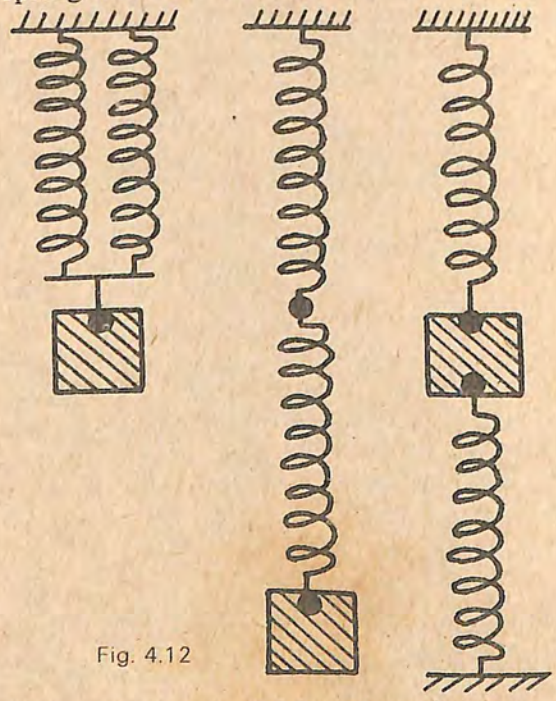


Fig. 4.12

4.7 A sphere is hung with a wire. 30° rotation of the sphere about the wire generates a restoring torque of 4.6 Nm. If the moment of inertia of the sphere is 0.082 kg. m^2, deduce the frequency of angular oscillations. (1.65 Hz)

4.8 A person normally weighing 60 kg stands on a platform which oscillates up and down harmonically at a frequency 2.0 s^{-1} and an amplitude 5.0 cm. If a machine on the platform gives the person's weight against time, deduce the maximum and minimum readings it will show. (Take g = 10 ms^{-2}.) (107 kg; 12 kg)

4.9 A cylindrical wooden block of cross-section 15.0 cm^2 and mass 230 g is floated over water with an extra weight 50 g attached to its bottom. The cylinder floats vertically. From the state of equilibrium it is slightly depressed and released. If the specific gravity of wood is 0.30 and g = 9.8 ms^{-2}, deduce the frequency of oscillation of the block.

(0.79 s^{-1})

UNIT 5

Wave Motion

In this Unit, we shall study some general facts about wave motion. For example, one important feature of wave motion is that no matter is transferred in the process, only energy is transferred. Another feature is that the velocity of waves relative to the medium depends only on the medium and not on the nature or motion of the source. These and other related matters we shall examine, including how waves may be represented mathematically and graphically, how energy is transferred, and so on.

5.1 Waves on Water and Strings

If we ever stood at a sea shore, we would need no description of waves. In a large tank, the waves are again seen, though less prominently. And in a small water basin, we could generate waves by dipping a finger or dropping a pebble (Fig. 5.1). All these are waves on the surface of water. It will be noted that there are alternate crests and troughs on the water surface and with the passage of time the crests and troughs travel along the surface of water. The whole set of crests and troughs is called *waves* and the material in which the waves travel (here, the water surface) is called the *medium*.

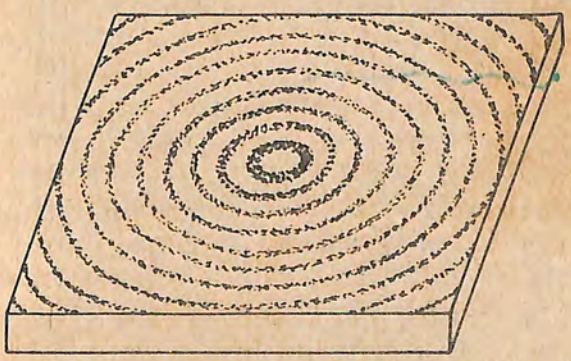


Fig. 5.1 Waves on the water surface in a basin. As time passes, the crests and troughs travel along the surface of the water.

Let us now consider a string held under some tension. If one end of it is suddenly shaken sideways, a kink or pulse starts off from there and travels along the string. In the process, each part of the string undergoes a disturbance about its own mean position and this disturbance occurs for the different parts in a particular sequence. In Fig. 5.2, the travelling of a pulse is shown by seven stages a, b..., g corresponding to time 0, t_1, $2t_1$... $6t_1$. The pulse which is at the left end at time 0 travels by a distance x_1 in each time interval t_1.

Let us have a closer examination of this case. First, we notice that no part of the string travels along with the pulse, i.e., no 'matter' is transferred. Only the information (here, the pulse) fed at one end is transferred along the string. In waves over the water surface also, the same is true. So, one important feature of wave motion is that *no matter is transferred* along the direction of wave transmission.

The velocity of wave transmission in a medium is called *wave velocity*. We shall denote it by c. For example, in Fig. 5.2, the pulse starting at $t = 0$ travels a distance x_1 in each time interval t_1, so its speed is x_1/t_1. Fig. 5.2 shows another pulse started at time $3t_1$. It is much weaker, but that also travels the same distance x_1 in each time interval t_1. A third pulse started at time $5t_1$ is of a

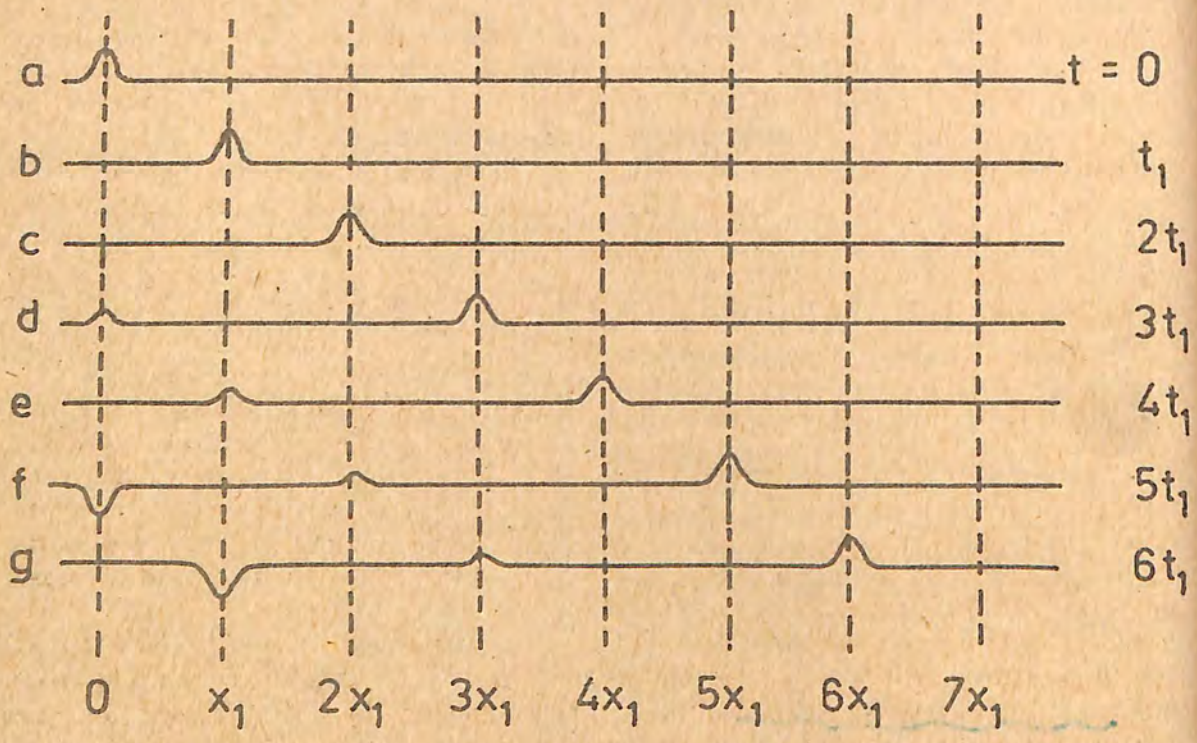


Fig. 5.2 The travelling pulses over a string a, b, c . . refer to the stages at successive intervals . . .

different shape; but that, too, travels at the same speed. So, the wave velocity does not depend on the size and shape of the pulse. A similar result holds for waves on the water surface and, in fact, for all kinds of waves. So, the second important feature of wave motion is that *the wave velocity relative to the medium depends only on the nature of the medium*, and it does not depend on the nature of the source, i.e., on the size or shape of the disturbance to be transmitted. (Some small variations do occur, but at present we shall omit that detail.) If the source is moving relative to the medium—like an aeroplane in air—then also the velocity of waves *relative to the medium* remains unchanged. This means that if the source velocity is V, then the wave velocity relative to the medium is still c in all directions, so that, relative to the source, the wave travels at a speed $c - V$ in the forward direction and $c + V$ in the direction behind the source. In contrast, consider a gun in an aeroplane which shoots bullets at a speed c'. Then, relative to the air, the bullet velocity will be $c' + V$ in the forward direction and $c' - V$ in the backward direction, whereas, relative to the source (gun), it is c' in all directions.

5.2 Sound Waves

When we are listening to a sitar recital, the sitar wire oscillates, the air as a medium carries these oscillations around, and when they reach our ears something in the eardrum carries the message to our brain. Thus, we hear. The air does not travel from the sitar to our ear; the oscillations (which have the message) are transmitted through the air. The velocity of this transmission depends only on the *medium* (air in this case) and does not depend* on whether the sound is loud

* For very high frequency waves, the velocity varies somewhat with frequency. This is called *dispersion*. But, presently, we omit such details.

or low, hoarse or shrill, music or speech.

For normal air, the velocity of waves is ~ 350 m/s, which means ~ 1200 km/hr. That is why the sound of the horn of an approaching car reaches us much faster than the car itself. The velocity of waves through the earth is about ten-fold. Compare these speeds with the usual speeds of wind. A speed of ~ 50 km/hr for wind is considered quite high.

Longitudinal and Transverse Waves

There is one distinction between the waves on water (or on strings) and the waves in air. If we desire to transmit the waves in air (or in any gas) in some direction x, then at the source we need to produce the disturbances along the same direction x. We cannot see air molecules. But we can observe the motion at the source (like a tabla membrane) or at the receiver (like a microphone membrane) by delicate instruments. We then find that the to-and-fro motion of the air molecules must be along the direction of wave motion. For this reason, the waves in air are called *longitudinal waves*. In contrast, the waves on strings are called *transverse*, because the to-and-fro motion of the particles about their mean positions are perpendicular (transverse) to the direction of wave motion. Longitudinal waves can also be transmitted through a liquid. In a long spring (slinky), we can create both transverse and longitudinal waves; a sudden compression along the length produces a longitudinal wave, while a sudden sideways kink produces a transverse wave (Fig. 5.3). We can do the same in a solid rod.

In general, in gases and liquids, only longitudinal waves can be transmitted. They do not have rigidity, but only volume elasticity. Therefore, only compressions and rarefactions travel in them, i.e., variations of pressure. For this reason, the longitudinal waves are also called *pressure waves*.

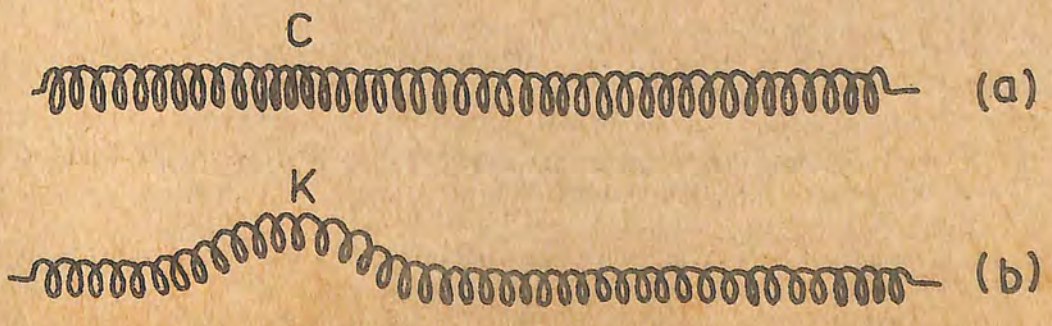
In contrast, solids have rigidity as well as volume elasticity. Therefore, in solids, both transverse waves (involving distortions) and longitudinal waves (involving compressions) can be transmitted. The speeds of the two kinds of waves are in general unequal.

A string under tension is a special situation. The string is too soft and shows negligible rigidity. The reason for wave transmission is the tension applied on the string. Thus, the velocity of wave transmission is not a property of the material of the string alone, but involves the external factor—tension—which can be adjusted. A sitar wire and a tabla membrane have this special feature. One often notices a musician adjusting the tension in these instruments, while no such thing exists, for instance, in a flute or a harmonium.

Electromagnetic Waves

All the waves so far discussed may be

Fig. 5.3 In a slinky, both longitudinal and transverse waves can be transmitted: (a) a compression C (b) a kink K.



grouped together as *mechanical waves* or *elastic waves*. There are entirely different kinds of waves which can be transmitted across a vacuum or free space, i.e., they do not need a material medium at all. They are called *electromagnetic waves*, because electrical and magnetic effects can be transmitted through a vacuum and in the case of these waves these two travel interlinked together. Waves of light, radio waves, microwaves, x-rays and gamma rays are all electromagnetic waves. These waves are transverse, i.e., the variations of electric and magnetic fields are perpendicular to the direction of travel of the waves.

5.3 Velocities of Different Waves

For transverse waves over a string under tension, the velocity of waves is given by

$$c = \sqrt{\frac{T}{m}} \qquad \ldots (5.1)$$

where T is tension on the string (N) and m is the mass per unit length of the string (kg/m). One may check dimensionally that $\sqrt{T/m}$ has the dimensions of velocity. Qualitatively, the facts that c increases with increasing T and decreases with increasing mass can be seen by passing pulses over strings.

For longitudinal waves in a medium, whether a gas, a liquid or a solid, the wave velocity is given by

$$c = \sqrt{\frac{E}{d}} \qquad \ldots (5.2)$$

where E is the volume elasticity (N/m^2) and d is the density (kg/m^3) of the medium. The density of a solid is much larger than that of a gas, but the elasticity is larger by a greater factor. Hence, waves in a solid travel faster than those in a gas.

The volume elasticity in a gas depends on whether the conditions are *adiabatic or isothermal*. In the isothermal case, E equals the pressure P of the gas; in the adiabatic case, E is γ times the pressure P, where γ is the ratio of the specific heat at constant pressure and at constant volume. Newton deduced the result (5.2) and used E = P to get for a gas,

$$c = \sqrt{\frac{P}{d}} \qquad \ldots (5.3)$$

But Laplace considered that the oscillations occur so fast that the heat generated in compression or absorbed during rarefaction is not transferred out or in. The changes are, therefore, adiabatic and, hence, one must use $E = \gamma P$ in the equation (5.2) to get

$$c = \sqrt{\frac{\gamma P}{d}} \qquad \ldots (5.4)$$

At a given temperature, P/d is constant for a gas. Hence, the wave velocity is independent of pressure. With changing temperature, P varies according to

$$d = \frac{do}{(1+\alpha\theta)}, \alpha = \frac{1}{273} \qquad \ldots (5.5)$$

where θ is the temperature in degrees Celsius. Substituting in the equation (5.4) and using the binomial theorem ($\alpha\theta \ll 1$), we get

$$c = c_o(1 + \frac{1}{2}\alpha\theta) \qquad \ldots (5.6)$$

EXAMPLE 5.1

For air at standard temperature and pressure, the density is 0.00129 g/cm^3. Deduce the velocity of longitudinal waves, using (i) Newton's formula, (ii) Laplace's formula.

Solution

$$P_o = 76 \times 13.6 \times 980 \times 10^{-1} \frac{N}{m^2}$$

$$= \left(\frac{\text{kg ms}^{-2}}{\text{m}^2} \right)$$

$$d_o = 1.29 \text{ kg/m}^3$$

From Newton's formula,

$$c = \sqrt{\frac{76 \times 13.6 \times 980 \times 10^{-1}}{1.29}}$$

$$\left(\frac{\text{kg m}^{-1}\text{s}^{-2}}{\text{kg m}^{-3}} \right)^{\frac{1}{2}}$$

$$= 280 \text{ ms}^{-1}$$

From Laplace's formula, since γ for air is 1.4

$$c = \sqrt{1.4 \times 280 \text{ ms}^{-1}}$$

$$= 330 \text{ ms}^{-1}.$$

EXAMPLE 5.2

For steel, $E = 2.9 \times 10^{11}$ N/m^2 and $d = 8 \times 10^3$ kg/m^3. Deduce the velocity of longitudinal waves.

Solution

$$c = \sqrt{\frac{E}{d}} = \sqrt{\frac{2.9 \times 10^{11}}{8 \times 10^3}}$$

$$= 6 \times 10^3 \text{ ms}^{-1}$$

5.4 Harmonic Waves

Let a wave be transmitted along the x-axis with velocity c. This means that whatever is happening at the origin at the time t will occur at the position x at the time $t + x/c$ or whatever is happening at x at the time t must have occurred at the origin at the time $t - x/c$. The latter statement may be written in the following form:

$$\xi(x, t) = \xi(0, t - x/c) \qquad \ldots (5.7)$$

If the displacement at the origin is given by

$$\xi(0, t) = A \cos 2\pi \frac{t}{T} \qquad \ldots (5.8)$$

for $\xi(x, t)$, we get

$$\xi(x, t) = A \cos \frac{2\pi}{T}(t - x/c) \qquad \ldots (5.9)$$

This is the equation for a harmonic wave travelling along the positive direction of the x-axis. It gives the displacement at any position x at any time t. It can be seen that at a given position (x fixed), the function ξ (x, t) repeats at time intervals of T; also at a given time (t fixed), the function repeats at space intervals of cT, which is called *wave length* λ. We can write the equation (5.9) as

$$\xi = A \cos 2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right) \qquad \ldots (5.10)$$

On the left side, just ξ is written as a matter of convenience. It is understood that it is a function of t and x. The function is periodic in time (interval T) and periodic in space (interval λ).

From the equation (5.10), the particle velocity $\dot{\xi}$ (as distinct from the wave velocity c) at a general position x at a general time t is given by

$$\dot{\xi} = \frac{d\xi}{dt} = - v_o \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right);$$

$$v_o = \frac{2\pi}{T} A \qquad \ldots (5.11)$$

We note that the particle velocity is a varying quantity, while the wave velocity c is a constant for a given medium. The time and space variation of the particle velocity is of the same nature as that of displacement, and the velocity amplitude v_o is $2\pi/T$ times the displacement amplitude A.

EXAMPLE 5.3

Consider the wave $\xi = 2.2 \cos (300t - 0.24x)$. If the units for ξ, t and x are mm, sec and m, respectively, deduce (i) the amplitude, (ii) the frequency, (iii) the wave velocity, and (iv) the amplitude of particle velocity.

Solution

(i) Since cosine is a pure number, the unit for 2.2 is the same as that for ξ. Hence, amplitude = 2.2 mm.

(ii) Comparing with the equation (5.10), $2\pi/T = 300\ s^{-1}$.

$$\text{Frequency } \nu = \frac{1}{T} = 48 \text{ Hz}.$$

(iii) Wave velocity $c = \frac{\lambda}{T} = \frac{2\pi/T}{2\pi/\lambda} = \frac{300\ s^{-1}}{0.24\ m^{-1}} = 1250\ ms^{-1}$.

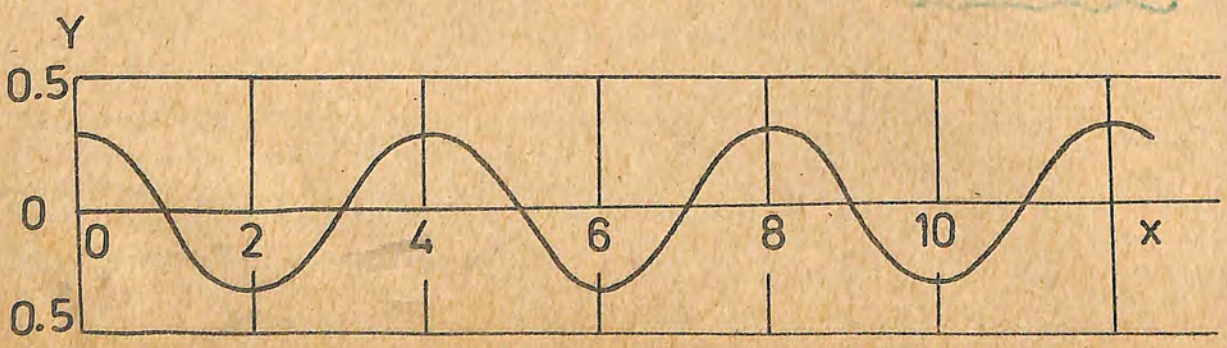
(iv) Amplitude of particle velocity $= \frac{2\pi}{T} A = 300 \times 2.2$ mm/s $= 0.66\ ms^{-1}$.

5.5 Graphical Representation of Wave Motion

Three parameters are involved in wave motion along a line: particle position x, time t and particle displacement ξ. In a two-dimensional graph, one can either plot ξ against t (at a given place x) or ξ against x (at a given time t). In Fig. 5.4 is shown ξ, x graph for a harmonic wave at a given time t. It will be noted that the scales for x and ξ are very different. Usually, in sound waves the displacement amplitudes are small fractions of a mm, whereas x extends to several metres. In a graph (unlike a photograph), convenient separate scales are to be chosen.

Another point about the graphical representation is that it can be used for both transverse and longitudinal waves. In longitudinal waves, the displacement ξ is in a direction parallel to that of wave motion, i.e., X-direction, yet the graph shows ξ in a direction perpendicular to X. The convention is that the displacements to the right are to be shown in the graph towards + Y axis, displacements to the left towards the −Y axis. In Fig. 5.5, the hollow circles in the first row represent the mean positions of a chain of equidistant particles in the medium. The arrows show (rather magnified) their longitudinal displacements at a given time. Obviously, the arrows are neither equal in length nor identical in direction. The solid circles in the second row show the instantaneous positions of the particles corresponding to the heads of arrows. Now for every arrow towards the right, we draw a proportionate line upwards and for every arrow towards the left, we draw a proportionate line downwards. Through the heads of these lines, a smooth curve is now drawn. This is the representation of the longitudinal displace-

Fig. 5.4 The (ξ, x) graph for a wave at a particular time. Scale X-axis; unit = 2 m, Y-axis, 1 unit = 0.5 mm. Note that although both ξ and x are lengths, the scales chosen are very different.



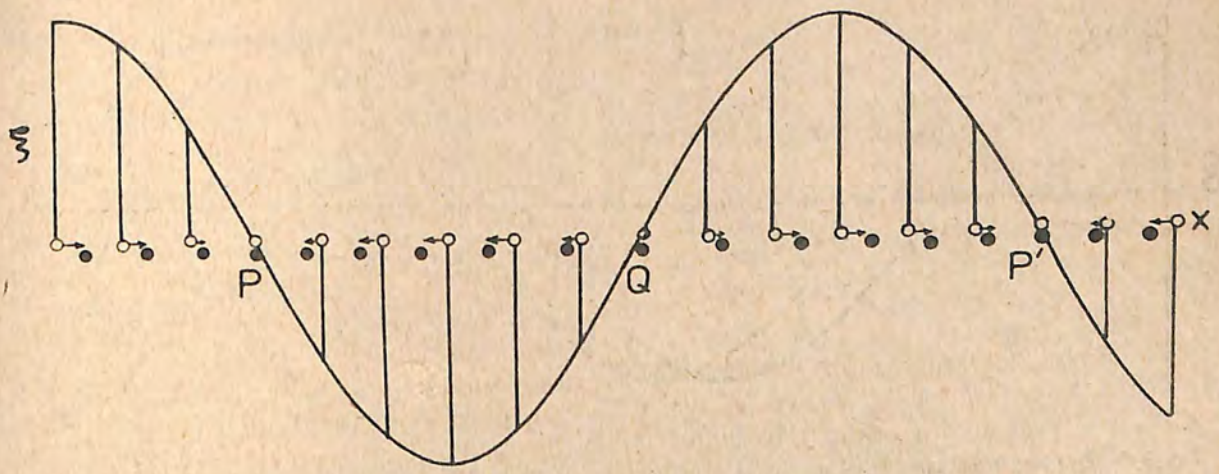


Fig. 5.5 Graphical representation of a longitudinal wave.

ments in the medium at the given time t. If we look at the solid dots, we note that around the positions P, P′ the particles have crowded together, while around the position Q they have separated farther than normal. Thus, there are alternate positions of maximum density (hence, maximum pressure) and minimum density. In longitudinal waves, alternate high and low pressures are transmitted with the waves.

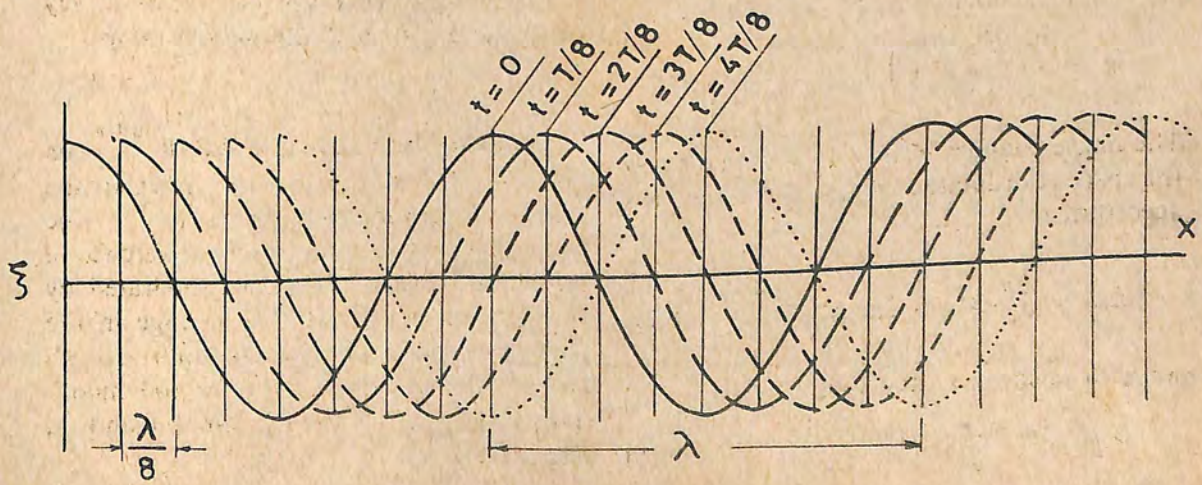
The travelling aspect of the wave can be shown drawing a series of ξ, x curves at successive intervals of time. In Fig. 5.6, this has been done for timings 0, T/8, 2T/8, 3T/8, etc. One can now visualize the crests and troughs moving ahead with time. In each time gap T/8, they move the distance $\lambda/8$ in the x direction.

In Fig. 5.7, two ξ, x curves are drawn for a small time interval $\triangle$ t. It is seen how the displacement changes $\triangle \xi$ in this interval vary with x.

5.6 Phase and Phase Difference

In any periodic motion, each of the quantities ξ, $\dot{\xi}$ and $\ddot{\xi}$ undergo a cycle of changes repeatedly, such that the stage at

Fig. 5.6 A travelling wave. The (ξ, x) graph is shown at several times with interval T/8 each. The wave form shifts continuously, but the particle at each position x only oscillates about the mean position.



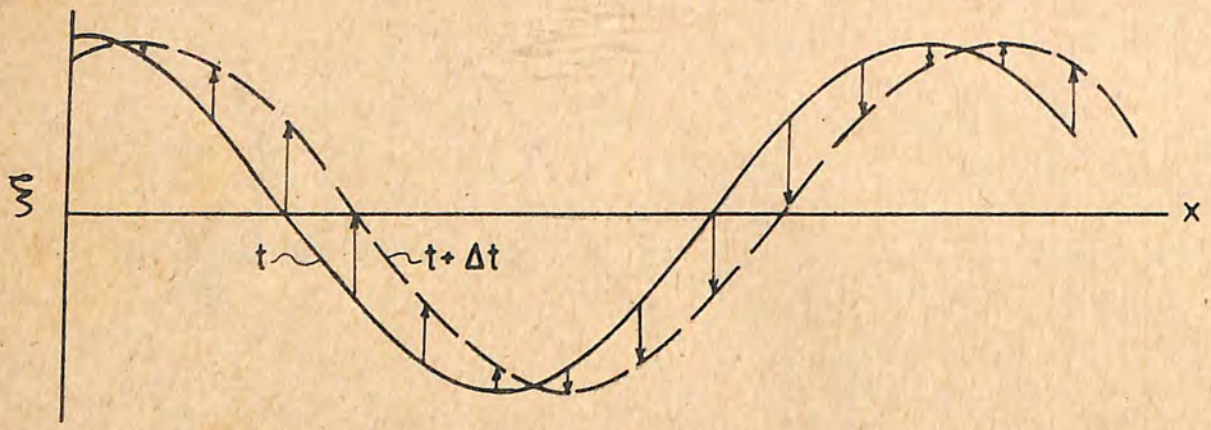


Fig. 5.7 Showing how two (ξ, x) curves at the time interval $\Delta\xi$ in displacements in the time Δt and hence are proportional to the particle velocity at each x at the given time.

the time t + T, t + 2 T,... t + nT is the same as at the time t. The different stages in a cycle may be described by such expressions as quarter-cycle, half-cycle, three-quarter-cycle, etc., from some chosen initial stage. However, a more useful way to specify the stage is to use the 'phase angle'. A harmonic wave travelling towards the +X-direction may be represented by

$$\xi = A \cos\left[2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right) + \phi_o\right] \quad \ldots (5.12)$$

The argument of the cosine is called the *phase angle* or simply *phase*, ϕ. Thus, the phase at x at the time t in a wave represented by the equation (5.12) is given by

$$\phi = 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right) + \phi_o \quad \ldots (5.13)$$

The phase changes both with the time t and the space coordinate x. With time, it changes according to

$$\Delta\phi = \frac{2\pi}{T}\Delta t = 2\pi\nu\Delta t \quad \ldots (5.14)$$

and with position x according to

$$\Delta\phi = -\frac{2\pi}{\lambda}\Delta x \quad \ldots (5.15)$$

The minus sign in the latter equation is to be noted. It means that in a wave moving towards the +X-direction the forward points are behind the phase, i.e., they reach the successive stages of vibration later.

One may use the 'phase' to define λ and T. In any wave, λ is the distance between points which have a phase difference 2π. Similarly, T is the time in which the phase at a given point changes by 2π. Very often, the vibrations which have a phase difference of integral multiples of 2π are said to be 'in same phase', and those which have a phase difference of multiples of π are said to be 'in opposite phase'. It is a convenient and meaningful language, but omits to say about the sequence of intermediate stages.

EXAMPLE 5.4

For the plane wave $\xi = 2.5 \cdot 10^{9.02\omega}$ cos (800 t − 0.82 x + π/2) write down (i) the general expression for phase ϕ, (ii) the phase at x = 0, t = 0, (iii) the difference of the phase between the points separated by 20 cm along x-axis, (iv) the change in the phase at a given place in 0.6 milli second, and (v) the amplitude at x = 100 metre. (take units of ξ, t, x as 10^{-5} cm, sec and m, respectively.)

Solution

(i) $\phi = 800\,t - 0.82\,x + \pi/2$. It does not depend on the fact that the amplitude here is decreasing with increasing x.

(ii) $\phi_o = \pi/2$, obviously.

(iii) $\Delta\phi = -0.82\,x = -0.82 \times .20$
$= -$ **0.164 radian**

(iv) $\Delta\phi = 800\,t = 800 \times 0.6 \times 10^{-3}$
$= 0.48$ radian.

(v) Amplitude at x is $2.5{\cdot}10^{-0{\cdot}02x} \times 10^{-5}$ cm, where x is to be in metre. At x = 100 m the amplitude is $A_{100} = 2.5{\cdot}10^{-2} \times 10^{-5}$ cm $= 0.025 \times 10^{-5}$ cm.

5.7 Wave Fronts

If we repeatedly dip our finger in water, a series of crests and troughs travel out. At any one instant, each crest has a circular form and each trough also has a circular form. If the wave is represented by a cosine function and an upward displacement is considered positive, then a given crest means phase $\phi = n{\cdot}2\pi$, (n = integer) and a given trough means $\phi = (m + \frac{1}{2})\,2\pi$, (m = integer). In other words, a given crest is the locus of constant phase $n{\cdot}2\pi$ and a given trough is also the locus of constant phase, $(m + \frac{1}{2})\,2\pi$. *The locus of points in a given phase is called a wave front.* In the case of water waves from a point source, the wave fronts have circular shape. In the case of waves in air from a point source, the wave fronts are of spherical shape. If we repeatedly dip a long straight bar in water, then at small distances the wave fronts will be straight lines. At a large distance from any source, the wave front always become very nearly plane.

For waves over the water surface from a point source, it will be observed that the heights of crests go on decreasing as we go to a larger radius r. This is because of the spread of the same energy over a larger wave front as r increases. Similarly, for waves in

Fig. 5.8 Waves get weaker as we go away from the source.

space, the energy spreads over larger spherical wave fronts and, hence, the amplitude falls as r increases.*

The equations (5.10) or (5.12) is for plane waves, by which we mean those waves in which the wave fronts are plane. In these waves, the phase at a given time depends only on x, so that the wave fronts are parallel to the yz plane. In such cases, if x_1, y_1, z_1 and x_2, y_2, z_2 are two points in space, the phase difference depends only on $x_1 - x_2$ and is $(2\pi/\lambda)(x_1 - x_2)$.

5.8 Energy Transmission in Waves

Wherever oscillations exist, there is oscillation energy, which changes between kinetic and potential forms. In a progressive wave, like that shown by the equation 5–10,

* In surface waves, the length of the wavefront is proportional to r; in space waves, the area of the wavefront is proportional to r^2. From these facts, one can see that amplitudes in the two cases are proportional to $1/\sqrt{r}$ and $1/r$, respectively.

the oscillations are transferred with the velocity c and, hence, energy is also transmitted. If ρ is the density of the medium and v_o the particle velocity amplitude, the energy per unit volume is

$$u = \frac{1}{2} \rho v_o^2 \qquad \ldots (5.16)$$

The energy transfer per unit area per second, called intensity I, is c times u, thus,

$$I = \frac{1}{2} \rho c v_o^2 \qquad \ldots (5.17)$$

If the displacement amplitude is A and the frequency ν, we have

$$v_o = 2\pi\nu A \qquad \ldots (5.18)$$

Thus, both u and I are proportional to (amplitude)^2 as well as (frequency)^2.

EXAMPLE 5.5

A 5 watt source sends out waves in air at frequency 1000s^{-1}. Deduce the intensity at a 100-metre distance, assuming spherical distribution. If c = 350 ms^{-1} and ρ = 1.3 kg/m^3, deduce the displacement amplitude.

Solution

Intensity at 100-metre distance:

$$I = \frac{5 \text{ watt}}{4\pi (100)^2 \text{ m}^2}$$

$$= 4 \times 10^{-5} \text{ watt m}^{-2}$$

The equation (5.17) with $v_o = 2\pi\nu A$ gives,

$$I = 2\pi^2 \nu^2 A^2 \rho c$$

$$= 2\pi^2 (1000)^2 A^2 (1.3)(350)$$

Equating the two values of I and solving for A,

$$A = 0.7 \times 10^{-7} \text{ m.}$$

(Note the extremely small magnitude of the displacement amplitude for sound waves in air.)

5.9 Reflection of Sound Waves

The commonest experience of sound reflection is in echoes heard in large halls and in the neighbourhood of hills. But one can also see ripples in a water basin being reflected by the walls or by any obstacles placed in between.

If the interval of time between the production of a sound and receiving an echo is t', the distance of the reflecting body is $t'/2c$, where c is the speed of the sound. This is the general principle of 'depth ranging' in lakes and oceans. A device called the *sonar* is used in navigation to locate under-water rocks, icebergs, whales, etc. The corresponding device in aircraft navigation is the *radar* using electromagnetic waves.

In nature, the most advanced use of sound reflection as a basis of navigation is made by bats, which do not have eyes. They send out high frequency sound waves, and from the reflections received they estimate not only the distance (from the time gap), but also the nature and size of the reflecting surface (from the intensity of the returned sound) and the direction (from the difference of the time gap for the two ears). Bats have their senses perfected to get all these results without the need of calculations.

In the case of light, reflection from smooth surfaces occurs according to laws well known to us. Whether the same laws hold for mechanical waves needs some examination. The best method is to study the reflection of ripples in a water basin. In a water basin (Fig. 5.9), put any straight obstacle AB. Now with the edge of a straight scale, produce a wave front like PQ and observe what happens, with the passage of time, to that part which goes towards AB. Successive parts of the crest strike AB and ultimately a new wave front P′Q′ is formed

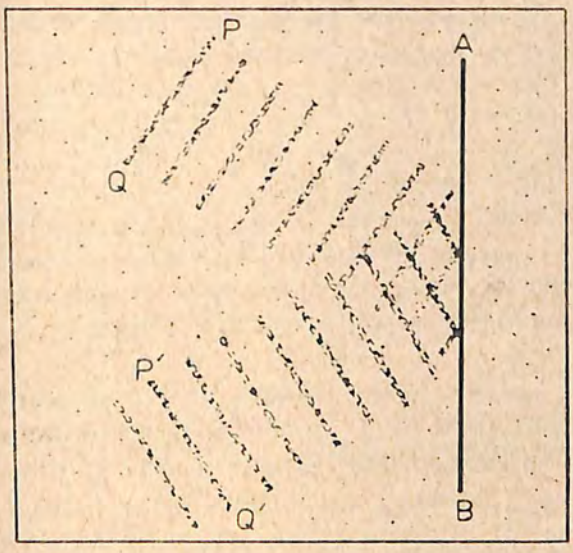


Fig. 5.9 The fate of a single wave front PQ at successive intervals is shown as it goes towards AB and gets reflected (schematic).

travelling away from AB*. (In Fig. 5.9, the intermediate stages are also shown.) Actually, the view is not as simple as shown in Fig. 5.9; the edges, in particular, are not distinct or sharp**. But the central part of the wave front gets reflected as shown. One can see experimentally that the angles made by P'Q' and PQ with AB are equal, but on the opposite sides of the normal to AB.

The ripples experiment can also be done, using a point dipper with concave reflectors. Fig. 5.10 shows what may be observed. If we use a single dip of S, then the different arcs represent the pulse (in part) at successive time intervals, ultimately converging at S' and then further expanding. In fact, a direct sound experiment on this principle makes a good demonstration. S would be a watch, and S' the position of a listener's ear. AB being a large concave surface ~ 1 m diameter. **One could deduce the focal length of the concave reflector by the formula $f = vu/(v + u)$, used in optics.**

5.10 Refraction of Sound Waves

When a wave reaches the boundary of two different media, it is found that it is partly reflected and party transmitted into the second medium.

To study this, we first experiment with

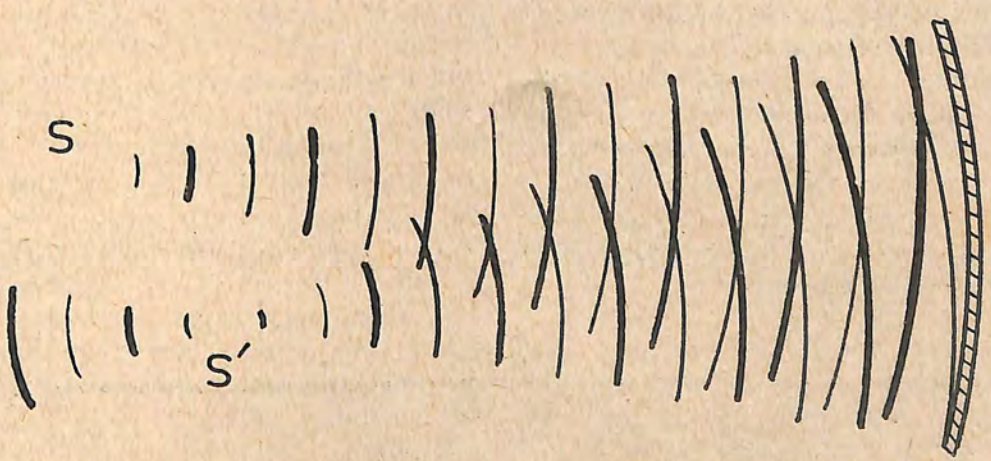


Fig. 5.10 Reflection of wave fronts developed from S by a concave reflector AB.

* The theoretical explanation why a given incident wavefront changes direction or shape in a particular way on reflection or refraction is given in Unit 7 on 'Optics' under the heading 'Huygen's construction'.

** Instead of a single crest, one can produce a series of them by a dipper rod oscillating at some frequency ν. In that case, if light is shined on the surface at the same frequency ν, one gets a 'stationary view'. One can then measure the wave-lengths in the incident and reflected waves, and also measure the angles of incidence and reflection with ease.

two strings joined together and kept under some tension. Fig. 5.11 shows AO and OB as the two strings joined together at O and kept under some tension. The velocity of the wave is c in AO and c′ in OB, where c′ < c because the string OB has larger mass per unit length. A kink is generated in the string AO. At the time t = 0, this kink has just reached the junction O; at t = t_1, we find a kink P′ in the string AO and another Q′ in the string OB; at t = t_2, we find these kinks travelliny to P″ and Q″, respectivelv.

We note, firstly, that reflected and transmitted pulses are created simultaneously. We note also that the distance Q′ Q″ is less than P′ P″, so that c′ < c. We note a third point also: while the kink Q′ is in the same direction (upward) as P, the reflected kink P′ is upside-down. We may repeat the experiment, initiating the incident kink in the string OB. In that case, the reflected kink is not upside-down. Experiments with several pairs of chords lead always to this common result.

(i) The wave coming from a medium, where the wave-velocity is larger, is reflected upside-down, which means with a phase change π.

(ii) The wave coming from a medium, where the wave-velocity is smaller, is reflected without the change of sign, i.e., with no phase change.

(iii) The wave transmitted into the second medium always goes without the change of sign.

For studying the change in the direction during refraction, we have to take at least a two-dimensional case. Ripples are convenient. There is a device for creating 'two' media: If the liquid depth d is small, the velocity of ripples depends on the depth – velocity decreasing as d decreases. Hence, if a part of the tank has a uniform disc immersed in it, the area thus covered has smaller liquid depth and hence smaller wave-velocity.

Now with a straight dipper, we can generate straight wave fronts in one medium. Fig. 5.12 schematically shows the wave fronts

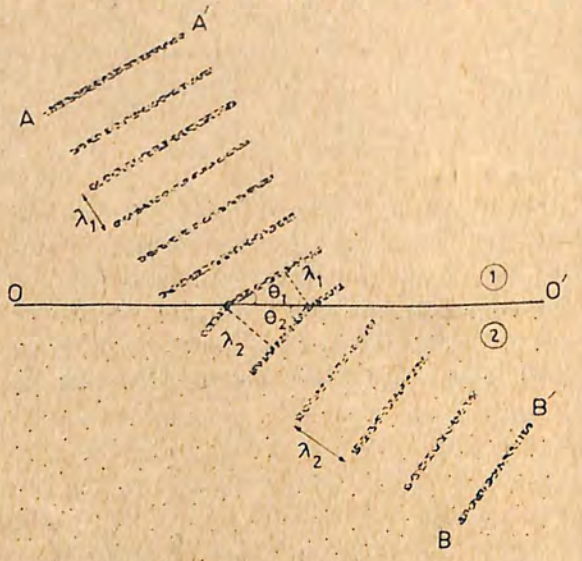


Fig. 5.12 Reflection of waves (schematic). The wave-velocity is larger in medium 2 than in 1. $\frac{C_2}{C_1} = \frac{\lambda_2}{\lambda_1}$

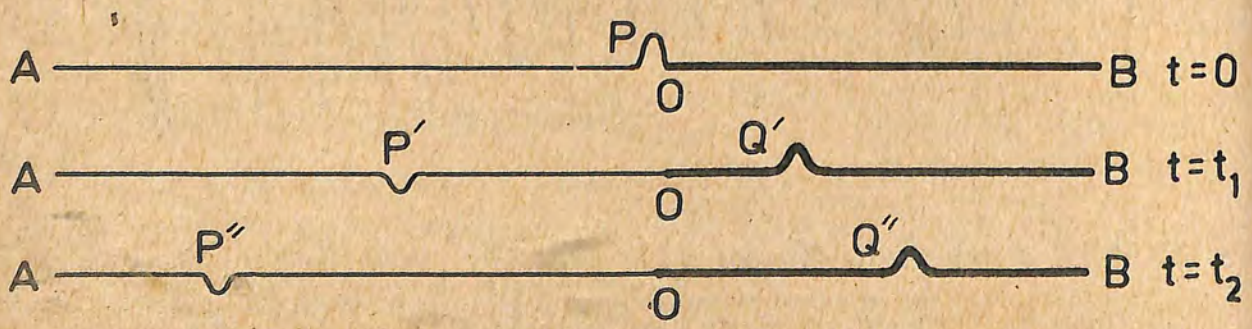


Fig. 5.11 Reflection and transmission of a kink at the junction of two strings.

like AA′ in the medium 1, and wave fronts like BB′ emerging in the medium 2. (The reflected ones are not shown.) One can in principle measure λ_1 and λ_2 so that one gets $c_1/c_2 = \lambda_1/\lambda_2$. If θ_1 and θ_2 are the angles of incidence and refraction by simple geometry, it can be seen that

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{\lambda_1}{\lambda_2} = \frac{c_1}{c_2} = \text{constant} \quad \ldots (5.16)$$

We, thus, experimentally see the wave fronts changing their direction and the wave lengths being unequal, and then understand why the ratio of sines is a constant for the given pair of media.

Thus, the law of refraction as given in light (called Snell's law) is seen to be obeyed by all waves in general. The wave theory gives the additional information that the refractive index must be equal to c_1/c_2.

A 'lens' for sound can be made, if between two circular plastic sheets sealed round the rim a gas other than air is filled*. About 1 m diameter and 30 cm bulge at the centre is suggested; the plastic sheet should be thin. Two persons may act, one as a 'source' and the other as a 'receiver'. As the receiver moves about, he hears a very loud sound in a particular region. One may verify that $(1/v) + (1/u) =$ constant for a given 'lens' in sound, as in light.

Dispersion

In mechanical waves, the wave-velocity does not change widely with frequency unless we go to very high frequencies. For this reason, it is not simple to show dispersion. For all practical purposes, we, therefore, ignore the dispersion and treat the wave-velocity c′ in a given medium as independent of frequency. For light, the 'colours' are related with wavelengths and hence with frequencies ($\nu = c/\lambda$). There the dispersion is easily shown by passing white light through a prism. Let us have a closer look at it. Experiments show that

$$\lambda_{\text{violet}} < \lambda_{\text{red}}; \quad \mu_{\text{violet}} > \mu_{\text{red}}$$

But we now see that

$$\mu = \frac{c\,(\text{in vacuum})}{c'\,(\text{in the medium})} \quad \ldots (5.17)$$

Therefore, in the case of light the dispersion is such that $d\mu/d\lambda$ is negative, hence, $dc'/d\lambda$ is positive.

5.11 Polarisation in Waves

Let us consider transverse waves over a string. If the length of the string is along the X-axis, **the transverse vibrations will be in the YZ plane. They could be along the Y- or Z-direction, or at any angle θ to the Y-**direction. If we have a wave in which all the directions in the transverse plane are equally present in the vibrations, the wave is called *unpolarised*. If, however, the vibrations are limited to just one direction in the yz plane, the wave is said to be polarised. In longitudinal waves, the disturbance ξ is along the line of transmission only; so no polarisation occurs.

If we pass a string through a vertical slit and make one end oscillate in various directions about the length, the slit would permit only the vertical vibrations to pass. It will cut off the horizontal vibrations.

We now describe a simple experiment with light which shows that light cannot be corpuscular but must have wave-nature, and the waves are transverse, not longitudinal.

If we see a source of light S (Fig. 5.13) through a polaroid A, about 50 per cent intensity is cut off. If the plate is rotated in its own plane by any angle, no further change is observed. Now, if a second polaroid B is introduced in the path, in one position of

* A higher density gas will be needed to produce a convex lens.

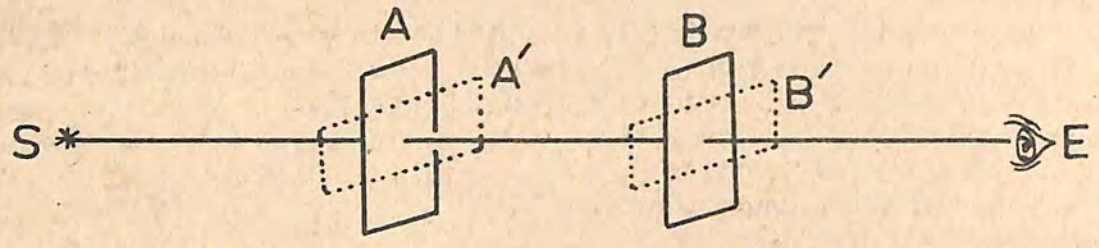


Fig. 5.13 An experiment with two tourmaline plates.

this plate all the light passed by A is passed through; if now B is rotated in its own plane, the transmitted light becomes weaker and weaker and it is cut off totally for rotation $\theta = 90°$ (position B′). With further rotation, the transmission improves, reaching the highest at $\theta = 180°$. Alternatively, we may keep B fixed and rotate the polaroid A. The same changes in the intensity are again observed*.

The experimental facts have no explanation if light is considered as particles (corpuscles) thrown by the source; it has, again, no explanation if light is considered as longitudinal wave motion. The only way of explaining the experimental facts is to consider light as *transverse wave motion*. In Fig. 5.14, we repeat the arrangement of Fig. 5.13, with the addition of x, y, z axes and the circles P, Q, R, S, T in which the transverse vibrations (in the yz plane, if x is the axis of transmission) are shown. At P (the source), the vibrations are equally distributed in the yz plane; before passing through the polaroid A, the vibrations are still equally distributed (see Q). But the polaroid A has some characteristic that it only allows vibrations in (say) the y-direction to pass and stops other vibrations. Since the displacements in the yz plane are vectors, each will split into y component and z component, out of which the first passes through and the second is stopped. Thus, after passing through the polaroid A, the light has vibrations along y only (see R). Now, the y vibrations travel up to the polaroid B′. This polaroid has the same nature as A, except that it is turned by 90° (see Fig. 5.13) and, hence, would permit only the z vibrations. Since the incident light (see S) has only y vibrations, the polaroid B′ passes nothing (see T). If B is turned by 90°, it is ready to pass y vibrations and since the incident light has only y vibrations, all of them pass**.

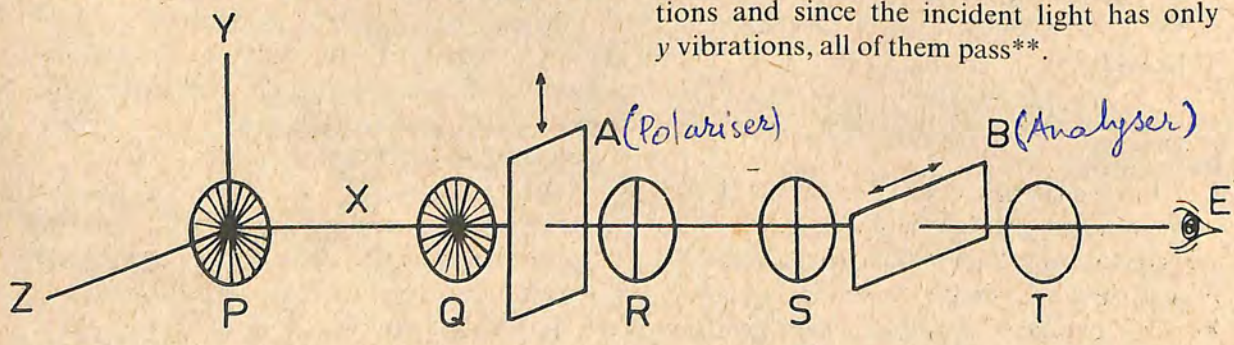


Fig. 5.14 Explanation of the two-polaroid experiment.

---

* If the transmitted intensity is quantitatively measured (say, with a photo-cell), we find that $I = I_0 \cos^2\theta$, where I_0 is the maximum intensity, and θ is the rotation of any one of the polaroids in its own plane relative to the maximum transmission setting.

** If the second polaroid is turned from B position by angle θ, then a $\cos\theta$ component of amplitude passes; therefore, a $\cos^2\theta$ part of the intensity. Thus, the experimental observations are quantitatively explained by the transverse wave model of light.

The light is said to be unpolarised if it has equal vibrations in all directions in a plane perpendicular to the line of transmission. If the vibrations are limited to just one direction in a plane perpendicular to the line of transmission, the light is said to be polarised or, rather, *plane polarised*. The first polaroid A is called the *polariser*; the second polaroid is called the *analyser*; but these are identical in behaviour and could interchange positions.

Observe the light reflected from a simple glass-sheet or water surface at an inclination $\sim 55°$ through a polaroid. If the polaroid is rotated in its own plane, the intensity changes quite widely. It shows that this reflected light is polarised.

In mechanical waves, the occasion to observe polarisation does not come easily. Sound waves in air are longitudinal. On strings one can always see polarisation, but its practical utility is small.

5.12 Doppler Effect

Doppler observed that when the source, the medium and the observer are in motion, the frequency of sound as received by the observer is different from the frequency of sound emitted by the source. This effect is called the *Doppler effect*.

The wavelength is defined by the separation of successive crests in the medium. If the source emits ν crests persecond (i.e. its actual frequency is ν), then the wavelength λ' is given by

$$\lambda' = \frac{c'}{\nu} \qquad \ldots(5.18)$$

where c' is the velocity of the waves *relation to the source*. If neither the source, nor the medium is in motion, then c' is just equal to the true speed of sound c'' in the medium. If the medium has speed V_m towards the observer, we get $c' = c + V_m$. If the source has speed V_s towards the observer, we get $c' = c - V_s$. If both the medium and the source are in motion, we get $c' = c + V_m - V_s$. Hence, for the general case

$$\lambda' = \frac{c'}{\nu} = \frac{c + V_m - V_s}{\nu} \qquad \ldots(5.19)$$

It is to be noted that, the change of wave length from $\lambda = c/V$ to the value λ' as given above is not apparent, but actual

The apparent frequency of sound, as experienced by the observer, is given by an equation like (5.18), written as :

$$\nu' = \frac{c''}{\lambda'} \qquad \ldots(5.20)$$

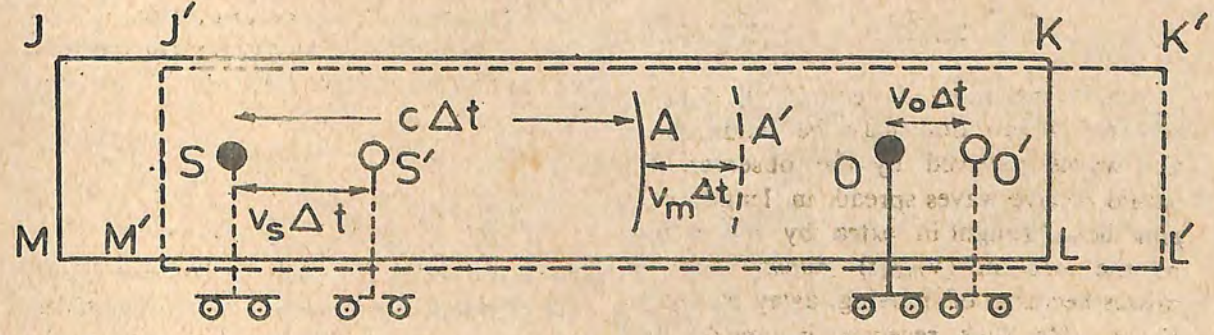


Fig. 5.15 Showing the motion $V_s\,\Delta t$, $V_0\,\Delta t$, $V_m\,\Delta t$ of the source, observer and medium, respectively, in the time Δt, and also the wave transmission by $c\Delta t$ in the medium in the same time.

* This does not apply to light, where relative velocities do not follow the simple relations discussed here. This will be studied under 'Relativity'.

where c'' is neither epual to c, nor c', it is the speed of sound waves *relative to the observer.* The λ' here is that which is given by eq. (5.19). Now, if the medium has a speed V_m towards the observer, we get $c = c + V_m$. If the observer has a speed V_o away from the source, we get $c'' = c - V_o$. If the medium and the observer both are in motion, $c'' = c + V_m - V_o$. For the general case therefore

$$\nu' = \frac{c''}{\lambda'} = \frac{c + V_m - V_o}{\lambda'} \qquad (5.21)$$

Substituting for λ' from eq. (5.19) we get

$$\nu'' = \nu \frac{c + V_m - V_o}{c + V_m - V_s} \qquad (5.22)$$

Fig. 5.15 gives a physical description of the situation. S is the source, O the observer and JKLM a representation of the medium. In a time Δ t, the wave travels in the medium by distance $SA = c\,\Delta t$. The medium itself travels in time Δ t by distance $AA' = V_m \Delta t$. Also the source travels in time Δ t by distance $SS' = V_s \Delta t$. Thus the crest emitted by the source at the *start* of the interval Δt reaches A', while the crest emitted at the *end* of this interval is emitted at S'. Thus $\nu \Delta t$ crests are spread over a length $S'A' = c\Delta t + V_m \Delta t - V_s \Delta t$. giving the wavelength λ' (separation of successive crests) as

$$\lambda = \frac{(c + V_m - V_s)\Delta t}{\nu \Delta t} = \frac{c + V_m - V_s}{\nu}$$

which is the same as eq. (5.19). Similar physical description may be made about the waves received by the observer. He would receive waves spread in length $c\Delta t$ *plus* those brought in extra by the motion of the medium ($V_m \Delta t$) *minus* those he misses because of moving away by length $V_o \Delta t$. He thus receives in time Δt that number of waves which are spread over length $c\Delta t + V_m \Delta t - V_o \Delta t$. This number is obtained by dividing this length by λ'; and then the number divided by Δt gives the apparent frequency ν' You may check that it leads to eq. (5.22).

The medium seldom has an appreciable velocity compared with the speed *c*. Hence, one could reduce the equation (5.22) to

$$\nu' = \nu \frac{c - V_o}{c - V_s} \qquad \ldots(5.23)$$

As an example of the Doppler effect, one may consider the sonar for the determination of the velocity of a submarine relative to a ship. In water $c = 1500\ ms^{-1}$, and if a submarine approaches the ship at a speed $5\ ms^{-1}$, it amounts to the 'source' approaching at $10\ ms^{-1}$ and, hence, the frequency ν' is greater than ν by 1 part in 150. For $\nu = 60{,}000\ s^{-1}$, the change is $400\ s^{-1}$. The sonar measures this difference and may be calibrated directly to give the approach velocity.

We have earlier described how a bat determines the distance and location of objects, using scattered sound waves. It also possibly finds the velocity of the objects around it by sensing the Doppler frequency shift.

In the case of the radar, electromagnetic waves are used, for which the theory does not apply in the form given, because these waves don't require a medium. But the measurements of time and distances depend on the velocities V_s and V_o and a result like the equation (5.23) comes out again. If V_s is the relative velocity of the source (relative to the observer), the equation (5.23) becomes (for $V_s/c \ll 1$).

$$\nu' = \nu \frac{c}{c - V_s} = \nu \left(1 + \frac{V_s}{c}\right) \qquad \ldots(5.24)$$

As in the sonar, the radar measures not only the distance and location of an aeroplane, but also its velocity by determining the frequency shift. V_s is to be taken as twice the approach velocity of the aeroplane because the 'source' is the image of the oscillator by the 'reflector' (the plane).

The Doppler effect is also used in astronomy. The light from stars when examined in a spectrograph gives several spectral lines. Compared with the corresponding spectral lines from a source on the earth, these lines are slightly shifted. This shift is attributed to the motion of the star along the line of sight. Generally, this shift is towards the 'red', which means a higher wave length, a lower frequency. It means that the stars are moving away from us.

EXAMPLE 5.6

A man standing near a rail line hears the whistle of an engine, which has the velocity 20 m/s. What frequencies does the man hear, if the true frequency of the whistle is 1000 Hz?

Solution

Given $V_o = 0$, and we assume $V_m = 0$. Hence,

$$\nu' = \nu \frac{c}{c - V_s}$$

We take the velocity of sound in air c = 340 m/s. When the engine is approaching, V_s is positive, + 20 m/s

$$\nu' = 1000 \frac{340}{340 - 20} = 1063 \text{ Hz.}$$

When the engine is going away, V_s is negative.

$$\nu = 1000 \frac{340}{340 - (-20)} = 944 \text{ Hz.}$$

EXAMPLE 5.7

A source and an observer are approaching one another with the relative velocity 40 m/s. If the true source frequency is 1200 Hz, deduce the observed frequency under the following conditions:

(i) All velocity is in the source alone,
(ii) All velocity is in the observer alone,
(iii) The source moves in air at 100 m/s towards the observer, but the observer also moves with the velocity V_o in the same direction.

Solution

For (i), $\nu' = 1200 \frac{340}{340 - 40}$

$= 1360$ Hz.

For (ii), $\nu' = 1200 \frac{340 + 40}{340}$

$= 1340$ Hz,

because $V_o = -40$ (opposite to the source → observer direction).

For (iii), $\nu' = 1200 \frac{340 - 60}{340 - 100}$

$= 1400$ Hz,

because $V_o = 100 - 40 = 60$ m/s.

Notice that although the relative velocity between the source and the observer is the same in all the three cases, ν' value depends on the absolute velocities of the source and the observer in the medium.

EXERCISES

5.1 (a) How do you infer from Fig. 5.2 that pulses of all shapes travel with equal speed?

(b) How much time does sound take to cross a hockeyfield (~ 100 m)?

(c) If an explosion takes place at the bottom of a lake, will the shock waves in the water be longitudinal or transverse?

(d) Audible frequencies have a range 40 Hz to 30,000 Hz. Express the range in terms of (i) period T, (ii) wave length λ in air, and (iii) angular frequency.

5.2 (a) A displacement wave is represented by

$$\xi = 0.25 \times 10^{-3} \sin (500\,t - 0.025\,x)$$

where ξ, t and x are in cm, sec and metre, respectively. Deduce (i) the amplitude, (ii) the period, (iii) the angular frequency, (iv) the wave length. Deduce also the amplitudes of particle velocity and particle acceleration.

$(0.25 \times 10^{-3}$ cm; $\pi/250$ s; 500 s^{-1}; $20\pi/25$ m; 0.125 cm/s; 62.5 $cm/s^2)$

(b) Two waves of angular frequencies 50 and 5000 rad/s have the same displacement amplitude 3×10^{-5} cm. Deduce the acceleration amplitudes for them.

$(7.5 \times 10^{-2}$ cm/s^2; 7.5×10^{2} $cm/s^2)$

5.3 (a) In the equation 5.10, use the sine in place of the cosine and deduce the corresponding expressions for $\dot{\xi}$ and $\ddot{\xi}$. Check if the statement about the phase relations of ξ, $\dot{\xi}$ and $\ddot{\xi}$ still holds true.

(b) Write the equation for a wave travelling towards the $-x$ axis. Check the statement about the phase relations of ξ, $\dot{\xi}$ and $\ddot{\xi}$.

5.4 A drop of water, 2 mm in diameter, falling from a height of 50 cm in a bucket generates sound which can be heard from the 5-metre distance. Take all the gravitational energy difference as going into the sound form, the transformation being spread in time over 0.2 second. Deduce the average intensity and the amplitude of vibration at the listener's end.

$(3.3 \times 10^{-7}$ W m^{-2}; 6×10^{-9} m)

5.5 A and B are two points on a water surface where waves are generated. What is the phase difference (a) if A and B are on the same wave front, but separated in distance by the length 5 λ, (b) if A and B are on successive crests, but linearly separated by the distance 3.5 λ, (c) if A and B are in successive troughs?

$(0; 2\pi; 2\pi)$

5.6 Show that the volume strain in any longitudinal wave is given by $-d\,\xi/dx$, where ξ is the displacement at the plane x. Hence, show that for a harmonic wave of amplitude A and wave-length λ in a medium of volume elasticity E, the excess pressure p is given by

$$p = EA\,\frac{2\pi}{\lambda} \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda}\right)$$

Deduce the pressure amplitude, if $E = 1.6 \times 10^{3}$ N/m^2, $A = 4 \times 10^{-7}$ m, $\lambda = 0.5$ m.

$(8 \times 10^{-3}$ $N/m^2)$

5.7 (a) Is it necessary that the amplitude be constant over a given wave front?

(b) Can two wave fronts belonging to a given wave system cross over?

5.8 For plane waves in air of the frequency 1000 Hz and the displacement amplitude 0.2×10^{-7} m, deduce (i) the velocity amplitude, (ii) the intensity (Take $\rho = 1.3$ kg/m^3, $c = 340$ ms^{-1})

$(1.3 \times 10^{-4}\ \text{ms}^{-1};\ 3.7 \times 10^{-6}\ \text{W m}^{-2})$

5.9 In the two diagrams below, PQ is the dividing surface between two media, and 1, 2 are the wave fronts at the time t_0 and $t_0 + t_1$, respectively. Deduce the wave fronts in the upper medium for the time $t_0 + 2\,t_1$ and $t_0 + 3\,t_1$ in each case.

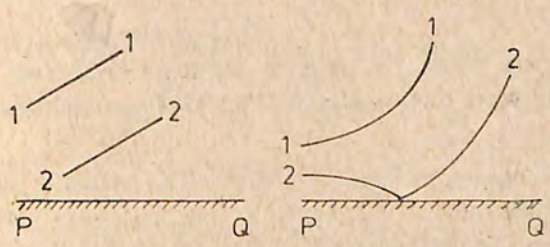


Fig. 5.16

5.10 Two polaroids are set to give the maximum transmission I_0. Deduce the transmission, if (i) the first polaroid is rotated by 30° clockwise, (ii) the second polaroid is now rotated by 30° clockwise, and (iii) the first is now rotated by 60° anti-clockwise.

(3/4; 1/4)

5.11 A stationary source emits sound of $\nu = 1200$ Hz. If wind blows at the speed 0.1 c, deduce (i) the percentage change in the wave length, (ii) the change in the frequency for a stationary observer on the wind-side of the source. Repeat the calculations for the case when there is no wind, but the observer moves at 0.1 c speed towards the source.

(10 percent increase; zero; zero; 10 percent increase)

5.12 Consider a source moving towards an observer at the speed $V_s = 0.95$ c. Deduce the observed frequency if the original is 500 Hz. (Think what would happen if $V_s > c$. Jet planes moving faster than sound are now so common?)

$(10^4\ \text{Hz})$

5.13 If c_1 is the root-mean-square thermal speed of molecules in a gas and c is the speed of sound waves in that gas, show that the ratio c/c_1 is independent of the temperature and is the same for all di-atomic gases.

5.14 For a string of mass 1.24 g/m, deduce the velocity of waves under tension (i) 10 N, (ii) 100 N.

$(90\ \text{ms}^{-1};\ 286\ \text{ms}^{-1})$

5.15 Show that for air the wave-velocity increases by nearly 0.6 ms^{-1} for each 1° C rise of the temperature.

5.16 For aluminium, the bulk modulus of elasticity is 7.5×10^{10} N/m^2 and density is 2.70×10^3 kg m^{-3}. Deduce the velocity of longitudinal waves in aluminium.

$(5 \times 10^3 \text{ ms}^{-1})$

5.17 Show that in a harmonic wave $\dot{\xi}$ is ahead of the phase by $\pi/2$ relative to ξ, and $\ddot{\xi}$ is further ahead of the phase by $\pi/2$.

5.18 Given that Avogadro number is 6×10^{26} per kg-mole and at normal temperature and pressure one kg-mole of a gas occupies 22.4 m^3 volume, deduce the average distance between the gas molecules at NTP. Compare this with the amplitude of displacement wave when the intensity is 1 watt/m^2 in air, taking effective frequency as 1000 Hz.

(Amplitude is $\sim 3 \times 10^3$ times the spacing)

5.19 The spectral line for a given element in light received from a distant star is shifted towards the longer wave length by 0.032%. Deduce the velocity of the star in the line of sight.

5·20 A radar wave has a frequency of 7.8×10^{9} s^{-1}. The reflected wave from an aeroplane shows a frequency difference of 2.7×10^4 s^{-1} on the higher side. Deduce the velocity of the aeroplane in the line of sight.

$(1.8 \times 10^3$ km/hr)

UNIT 6

Superposition of Waves

If more waves than one reach any part of space, their 'effects' add together. The principle of *superposition* says that if ξ_1, ξ_2, ξ_3... are the displacement vectors due to the waves 1, 2, 3... acting separately, then the displacement caused, when all the waves act together, is given by the vector sum

$$\xi = \xi_1 + \xi_2 + \xi_3 + \ldots \qquad \ldots (6.1)$$

It is to be noted that the 'effects' added are not intensities but the instantaneous displacements.

For simplicity, we shall deal only with cases where two waves superpose. Also, the displacement components will be taken to be along the same line. Then

$$\xi = \xi_1 + \xi_2 \qquad \ldots (6.2)$$

Each of ξ_1 and ξ_2 will be a function of time and space, so that ξ also will be a function of time and space. Among the cases of importance are the following:

(i) Two waves of the same frequency moving in the same direction (Interference of waves).

(ii) Two waves of the same frequency moving in the opposite direction (Stationary waves).

(iii) Two waves of slightly different frequencies moving in the same direction (Beats).

These cases will be taken up in this Unit, along with certain general applications.

6.1 Interference of Waves

A stethoscope is used by a doctor to hear the sound of heartbeats of a patient. The sound waves are transmitted to the ears through two tubes, which are of equal length. Quincke's tube device (Fig. 6.1) is different from this only in the sense that the two paths APB and AQB are not equal. In one U-shaped tube APB, there are openings at A and B; the other U-shaped tube can slide over the first, so that the path difference p = AQB − APB can be changed at will.

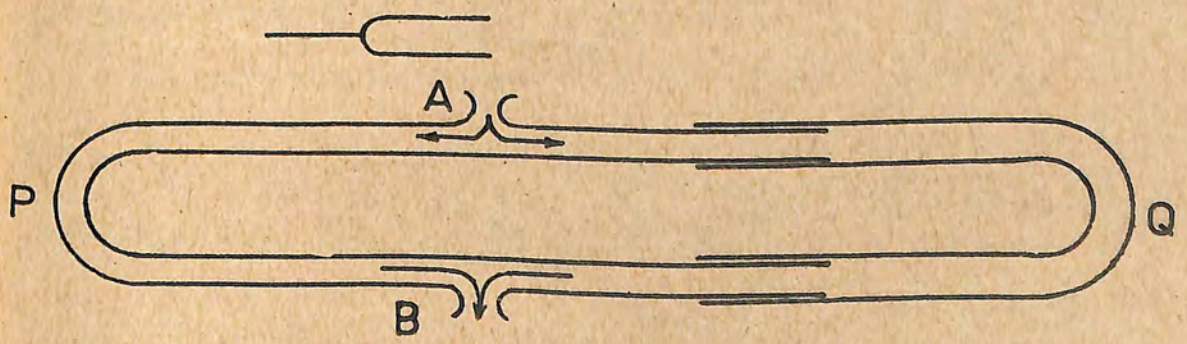


Fig. 6.1 Quincke's Tube.

Now, if a tuning fork is sounded near the opening A and a listener's ear is kept near B, it is found that the sound at B may be large or very small depending on the path difference p.

At A, the waves entering the two U-tubes are in the same phase. But at B, they will have a relative phase difference depending on p. If $p = m\lambda$, (m integer), the phase difference is $2m\pi$ and the equation (6.2) gives:

$$\xi = a_1 \cos \omega t + a_2 \cos (\omega t + 2m\pi)$$

$$\text{or } \xi = (a_1 + a_2) \cos \omega t. \qquad \ldots (6.3a)$$

Thus, the amplitude becomes $(a_1 + a_2)$. On the other hand, if

$p = (m + 1/2)\,\lambda$, (m integer), then the equation (6.2) gives

$$\xi = a_1 \cos \omega t + a_2 \cos \{\omega t + (2m + 1)\pi\}$$

$$\text{or } \xi = (a_1 - a_2) \cos \omega t \qquad \ldots (6.3b)$$

The amplitude now becomes $a_1 - a_2$. Since the intensity of the sound is proportional to the square of the amplitude, we get

$$I_{max} \propto (a_1 + a_2)^2; \quad I_{min} \propto (a_1 - a_2)^2$$

In general, for the path difference p, the phase difference is $(2\pi/\lambda)\,p = \phi$. The intensity is then

$$I \propto a_1^2 + a_2^2 + 2a_1 a_2 \cos \phi \qquad \ldots (6.4)$$

But we shall not give a proof of it here.

This phenomenon of two waves of the same frequency superposing to create conditions of variations of intensity is called the *interference of waves*. The maxima occur, if the path difference between the two waves is $m\lambda$, and the minima if it is $(m + \frac{1}{2})\,\lambda$.

Interference in Space

In the Quincke's tube, the waves were guided to the point B by the two U-tubes. But consider Fig 6.2 where S_1 and S_2 could be two dippers in a shallow water basin (like a ripple tank). At any instant of time, the crests and troughs produced by S_1 and S_2 are shown by full-line and dotted-line curves, respectively. Over the entire surface of water, we now have waves reaching from both S_1 and S_2 and, hence, interference is observed on the entire surface.

If S_1 and S_2 were two tuning forks (of the same frequency), then the waves from both would reach everywhere in the three-dimensional space, and interference would be observed in space.

For convenience, we shall discuss ripples. Wherever a crest meets another crest, there will be a stronger crest; wherever a trough meets another trough, there will be a stronger trough; and wherever a crest meets a trough, the effects will cancel out. In Fig. 6.2, the points of these three kinds are marked ·, o and □, respectively. With the passage of time, a point like P_1 has alternately high crests and high troughs, i.e., the disturbance there is strong. In contrast, at points like P_2 the state □, representing the cancellation of the crest and the trough, persists all the time, i.e., the disturbance there is very weak or zero*.

The positions of maxima and minima are given by a simple rule. If P is a general observation point, the quantity

$$p = S_2 P - S_1 P \qquad \ldots (6.5)$$

is called the path-difference. P has maximum intensity, if

$$p = 0, \lambda, 2\lambda, \ldots m\lambda \qquad \ldots (6.6a)$$

* If intermittent light of the same frequency as the dippers is shined on the ripples and the image is projected on a screen, we get a 'stationary view' of the ripples.

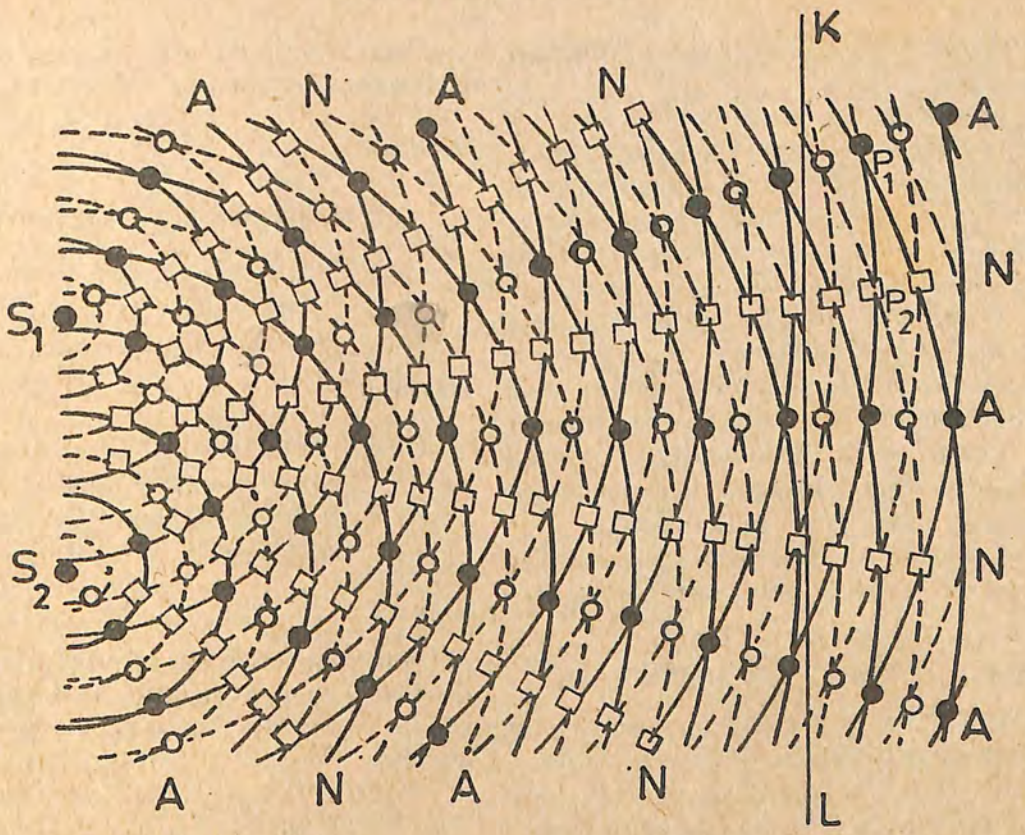


Fig. 6.2 Interference due to two sources S_1, S_2.

and minimum intensity, if

$$p = \frac{1}{2}\lambda, \frac{3}{2}\lambda, \frac{5}{2}\lambda \ldots \left(m + \frac{1}{2}\right)\lambda \quad \ldots (6.6b)$$

This assumes that the sources S_1 and S_2 send out waves in the same phase. If the sources give out waves with the phase difference π, the conditions for the maxima and the minima interchange. In general if S_2 is ahead of the phase relative to S_1 by ϕ_o, the conditions for maxima and minima become

$$\phi_o + (2\pi/\lambda)\, p = 2m\pi \text{ for maxima}$$
$$= (2m + 1)\,\pi \text{ for minima} \quad \ldots (6.7)$$

EXAMPLE 6.1

If sources S_1 and S_2 in Fig. 6.2 are two vertical radio antenna sending out waves of $\lambda = 80$ m and of equal intensity and the separation $S_1 S_2$ is 40 m, discuss the intensity of the resultant waves for points along $S_1 S_2$, along $S_2 S_1$ and along the perpendicular bisector of $S_1 S_2$ when (i) S_1 and S_2 are in the phase, (ii) S_1 and S_2 are in the opposite phase.

Solution

In case (i), the only phase difference is due to the differences of the path $S_2P - S_1P$. Along the perpendicular bisector, the difference is zero, hence, the maximum amplitude 2 a and the maximum intensity 4 I_1 where a_1 and I_1 are amplitude and intensity, respectively, due to a single antenna. Along $S_1 S_2$, the path difference $S_2P - S_1P = 40$ m $= -\lambda/2$. Along $S_2 S_1$, the path difference is $+\lambda/2$. Hence, the minimum amplitude

and the intensity will be 0 along both these directions.

In case (ii), let S_1 be ahead of the phase by π relative to S_2. Then, along the perpendicular bisector, $S_2P - S_1P = 0$ and the only phase difference π is due to the sources. Hence, in this direction the intensity will be zero. Along $S_1 S_2$, we have $S_2P - S_1P = -\lambda/2$, so the wave from S_1 lags in phase by π due to this path difference. As the source S_1 is ahead of the phase by π, the net result is that along $S_1 S_2$, the waves reach in the same phase with the intensity $4I_1$. Along $S_2 S_1$, the phase difference is π due to the source $+ \pi$ due to the path difference $= 2\pi$ in all. The waves, therefore, again reach in the proper phase, with the intensity $4I_1$.

Intensities at Maxima and Minima

If I_1 and I_2 are the intensities at the observation point due to the separate sources S_1 and S_2, then the corresponding wave amplitudes a_1 and a_2 are related through

$$I_1/I_2 = a_1^2/a_2^2. \qquad \ldots (6.8)$$

With both the sources acting, the amplitudes at the maxima and minima are $a_1 + a_2$ and $a_1 - a_2$, respectively. Hence, I_{max} and I_{min}, intensities at the maxima and the minima, can be written as:

$$\frac{I_{max}}{I_{min}} = \frac{(a_1 + a_2)^2}{(a_1 - a_2)^2}. \qquad \ldots (6.9)$$

If we use $r = a_1/a_2$ as the amplitude ratio, we can put the last result as:

$$\frac{I_{max}}{I_{min}} = \frac{(r + 1)^2}{(r - 1)^2},$$

$$\text{where } r = \sqrt{I_1/I_2}. \qquad \ldots (6.10)$$

EXAMPLE 6.2

If two sources of intensity ratio 100 : 1 interfere, deduce the ratio of intensities at the maxima and minima in interference.

Solution

$$\text{Amplitude ratio } r = \sqrt{100/1} = 10$$

$$\frac{I_{max}}{I_{min}} = \frac{(10 + 1)^2}{(10 - 1)^2} = \frac{121}{81} \simeq 3/2$$

EXAMPLE 6.3

In the interference with two sources of intensities I_o and $4I_o$, deduce the intensities at points where the phase differences are (i) 0, (ii) $\pi/2$, (iii) π, and (iv) $3\pi/2$.

Solution

Since the amplitude is proportional to the square root of intensity, the amplitudes will be A_o and $2A_o$. Now, in all the cases, we have

$$\xi = A_o \cos \omega t + 2A_o \cos (\omega t + \phi)$$

For case (i), $\phi = 0$. Hence,

$$\xi = (A_o + 2A_o) \cos \omega t.$$

$$\text{Amplitude} = 3A_o; \text{ Intensity} = 9I_o.$$

For case (ii), $\phi = \pi/2$. Hence,

$$\xi = A_o \cos \omega t - 2A_o \sin \omega t$$

$$= \sqrt{A_o^2 + 4A_o^2} \cos (\omega t + \delta),$$

$$\tan \delta = \frac{2A_o}{A_o} = 2$$

$$\text{Amplitude} = A_o \sqrt{5}; \text{ Intensity} = 5I_o.$$

For case (iii), $\phi = \pi$. Hence

$$\xi = (A_o - 2A_o) \cos \omega t.$$

$$\text{Amplitude} = -A_o,$$

$$\text{Intensity} = I_o$$

For case (iv), $\phi = 3\pi/2$. Hence,

$$\xi = A_o \cos \omega t + 2A_o \sin \omega t.$$

$$= \sqrt{A_o^2 + 4A_o^2} \cos (\omega t + \delta),$$

$$\tan \delta = -2$$

$$\text{Amplitude} = A_o\sqrt{5},$$

$$\text{Intensity} = 5I_o.$$

Interference Fringes and Fringe Width

If in Fig. 6.2 the sources S_1, S_2 are two flutes sounding the same tone, then in walking across a line like KL one alternately passes through the maxima and minima of sound. On the other hand, if S_1 and S_2 are the sources* of light of the same frequency, a screen like KL would show alternate bright and dark bands. These are called *fringes*, and the separation between the two successive maxima and minima is called *fringe width*, w. Since light has very small wave length, the separation S_1 S_2 has to be much less than the distance of the observation plane.

An expression for fringe width may be deduced from Fig. 6.3. S_1 and S_2 are the sources at the separation d, and KL is the observation plane at the distance D. O is a point equidistant from S_1 and S_2, and the observation point P is at a distance x from O. Now

$$S_1P^2 = D^2 + \left(x - \frac{d}{2}\right)^2$$

$$= D^2\left[1 + \frac{(x - d/2)^2}{D^2}\right]$$

Taking the square root and using the binomial theorem ($x \ll D$),

$$S_1P = D + \frac{1}{2}\frac{(x - d/2)^2}{D}$$

Similarly,

$$S_2P = D + \frac{1}{2}\frac{(x + d/2)^2}{D}$$

Hence, the path difference is given by

$$p = S_2P - S_1P = \frac{xd}{D} \qquad \ldots (6.11)$$

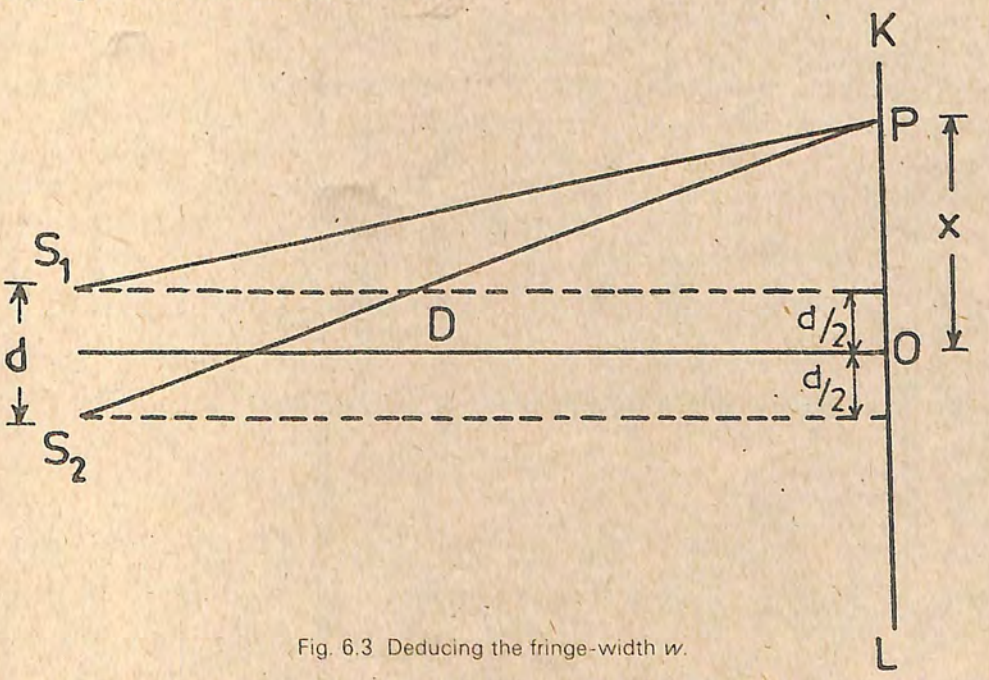


Fig. 6.3 Deducing the fringe-width *w*.

* In the case of light, a special arrangement is needed to keep the phase difference of S_1 and S_2 constant. We shall take it up in Unit 7.

Substitution in the equations (6.6) and (6.7) gives

$$x \text{ (for maxima)} = m \frac{D\lambda}{d} \qquad \ldots (6.12a)$$

$$x \text{ (for minima)} = \left(m + \frac{1}{2}\right) \frac{D\lambda}{d}. \qquad \ldots (6.12b)$$

The fringe-width w is the change of x when m changes to $m + 1$. Hence,

$$w = \frac{D\lambda}{d} \qquad \ldots (6.13)$$

A notable point is that if the sources S_1 and S_2 have a phase difference then the equation (6.12) will not have m as an integer, yet the equation (6.13) will always hold. For this reason, the equation (6.13) is used to deduce λ by measuring w, d and D.

6.2 Diffraction of Waves

In a ripple tank, let us put a stopper strip KL with an opening slit S (Fig. 6.4a). If straight wave fronts are generated on the left (say, by repeated dipping of a straight ruler), the observations on the other side of KL are interesting. If the width of the opening S is less than the wave length λ, we find that the ripples spread in all directions from S. If the width is larger than λ, then the ripples spread up to some angle from the normal direction, then become absent in a region like NN in Fig. 6.4b, and then again appear, though weakly.

This spreading of ripples around obsta-

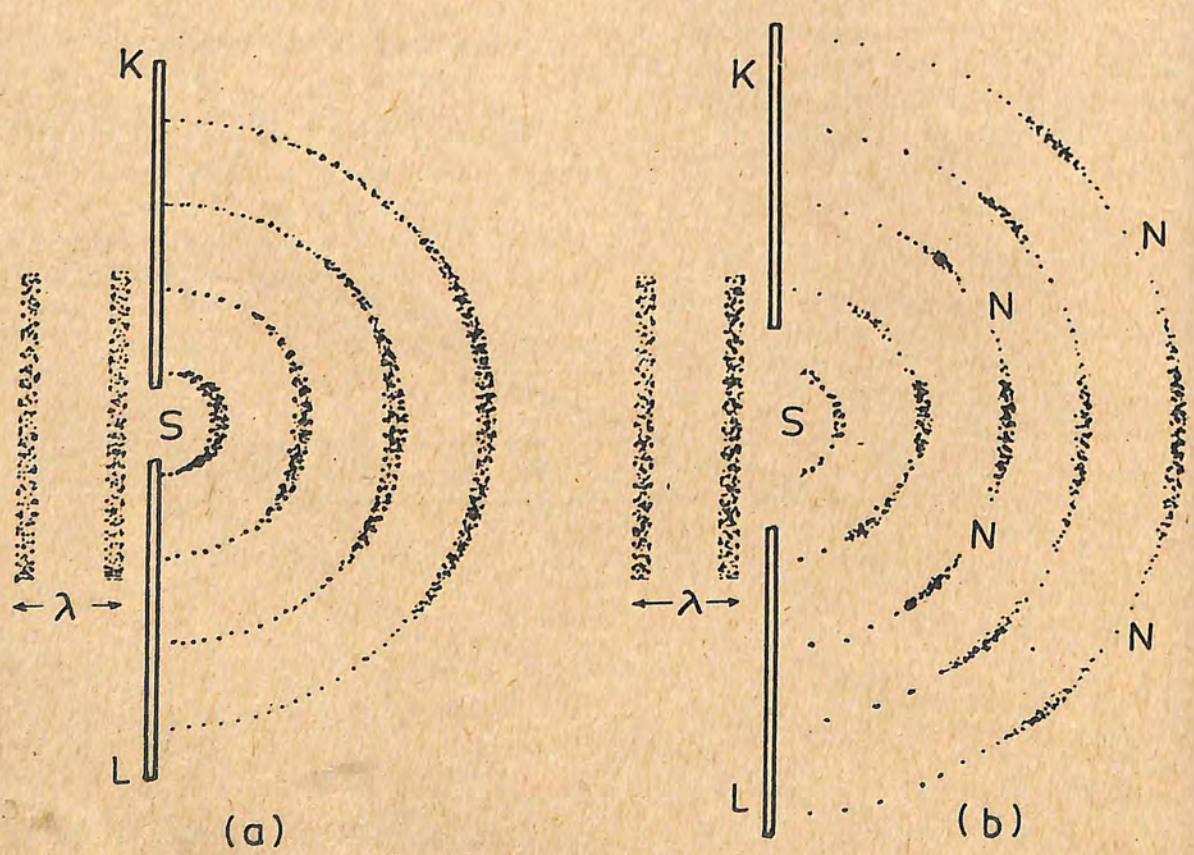


Fig. 6.4 Ripples passing through an opening S. Case (a) opening $< \lambda$, case (b) opening $\sim 2\lambda$

cles is called *diffraction*. Sound waves in air or anywhere show the same phenomenon of bending round corners, i.e., showing diffraction. It is a characteristic property in wave motion.

In sound, we take diffraction for granted because of the everyday experience of hearing the talks in a room even when standing well aside from an open window. But in the case of light, diffraction is not a common experience because seeing a speaker is not possible if one stands aside from a window. But the diffraction phenomenon is observed in light too, when the opening is quite narrow, say 10^{-3} cm or less in width. As the width decreases further, the diffraction of light is observed more clearly. From this fact, we draw two conclusions: (i) light, like sound, also shows wave behaviour, (ii) the wave length of visible light is very small relative to ordinary windows; it is certainly less than 10^{-3} cm. We shall study more about diffraction in the next Unit (Optics).

6.3 Beats

If two sources of slightly different frequencies send waves simultaneously, then at each point in space there is variation of intensity *with time*. Unlike interference, where intensity varies with position (not with time), here the intensity at a given place varies with time. Alternate loud and low sounds are heard. This phenomenon is called *beats*. The time from each loud sound to the next loud sound is called *one beat-period* and the number of such repetitions per second is called *beat-frequency*.

Mathematically, the explanation of beats is as follows:

Let the oscillations at the observation point due to the two sources be written as

$$\xi_1 = a_1 \cos 2\pi \nu t. \qquad \ldots (6.14)$$

$$\xi_2 = a_2 \cos 2\pi (\nu + \nu') t \qquad \ldots (6.15)$$

Here, the frequencies are ν and ν' and we take $\nu' \ll \nu$, i.e., the frequency difference is taken small. For the *phase* difference, we get

$$\phi = 2\pi (\nu + \nu') t - 2\pi \nu t$$
$$= 2\pi \nu' t \qquad \ldots (6.16)$$

i.e., the phase difference changes with time. Now, the resultant disturbance $\xi = \xi_1 + \xi_2$ is maximum when $\phi = 0, 2\pi, 4\pi \ldots$ and minimum when $\phi = \pi, 3\pi, 5\pi \ldots$. Hence, with the passage of time, we get alternate maxima and minima of sound. For each change of 2π in the phase difference, we get one beat. In one second, ϕ changes by $\nu' . 2\pi$ and, hence, we get ν' beats. Thus, the number of beats per second (beat-frequency) is equal to the difference of the frequencies of the two sources.

Graphically, one may understand beats through Fig. 6.5. The full-line curve is ξ, t graph for one oscillation, the dotted-line curve is for the second oscillation of a little larger period (lower frequency). At the

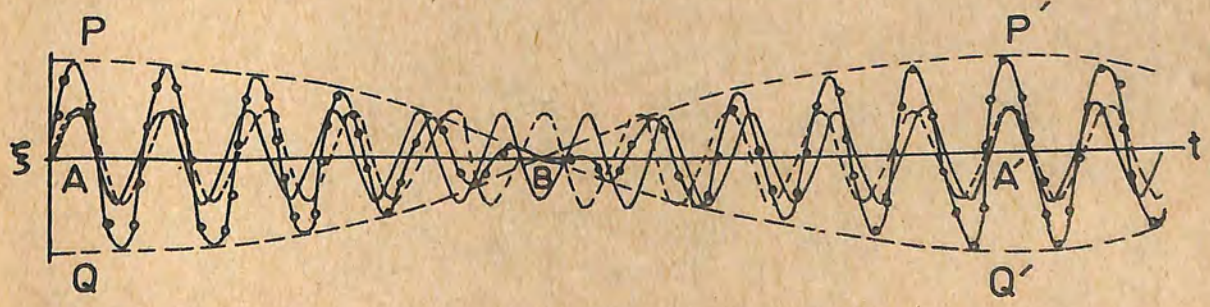


Fig. 6.5 Superposition of two waves of slightly different frequencies. The resultant amplitude varies with time.

start (A), the two oscillations agree in the phase. But as time advances, they go out of the phase (B), and, later on, come again in the phase (A'), and so on. The resultant oscillation is shown by the beaded curve and one notes the rise and fall of the resultant amplitude with the passage of time. In the figure, the curves PBQ′ and QBP′ show exactly how the amplitude varies with time. P to P′ is one beat-period, if one wave completes x oscillations in this time, the other wave completes $x + 1$ oscillations in the same time.

Musicians make good use of beats in matching the frequencies of two instruments. If there is a small frequency difference, it cannot be judged directly from the nature of the sound, but if the instruments are struck simultaneously, beats will be heard. Then one instrument is adjusted till the beats disappear.

In electronics, beat-frequency is often used. It is difficult to make low frequency oscillators, and, hence, the practice is to make two high frequency oscillators of small difference of frequency. Fig. 6.6 represents such beat-frequency oscillations, whose frequency is $1/T$, where T is the beat period. The beat principle is also used for frequency determination with accuracy. The oscillation of the unknown frequency is mixed with that of a standard oscillator and the latter is adjusted till the beats disappear.

EXAMPLE 6.4

A tuning fork P produces 5 beats/sec with another tuning fork Q whose frequency is known to be 284 Hz. Deduce the frequency of the fork P.

Solution

Frequency difference = beats/sec. = 5 Hz. The frequency of P is either 284 + 5 Hz or 284 − 5 Hz.

NOTE: To choose between the two answers, one slightly loads the fork P (with some wax). The frequency would decrease. Now, if the beat frequency decreases, the answer is 289 Hz; otherwise, it is 279 Hz.

6.4 Stationary Waves

We shall now take up the superposition of two waves of equal frequencies and equal amplitudes travelling through the same medium in *opposite directions*. The resultant effect is a wave which does not travel either way with time. For this reason, these waves are called *stationary waves*. In contrast, the waves so far discussed are called *travelling waves* or *progressive waves* because they travel along with some speed c.

The Graphical Method

One method of seeing the result of superposition of waves travelling oppositely is to write out the equations of the two waves and add them. But we shall use the graphical method of drawing the ξ, x graphs for the component waves at several equal time

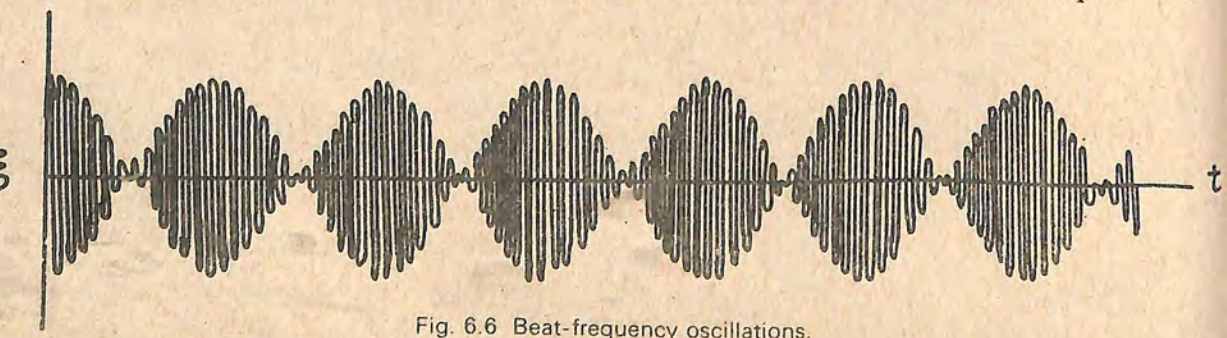


Fig. 6.6 Beat-frequency oscillations.

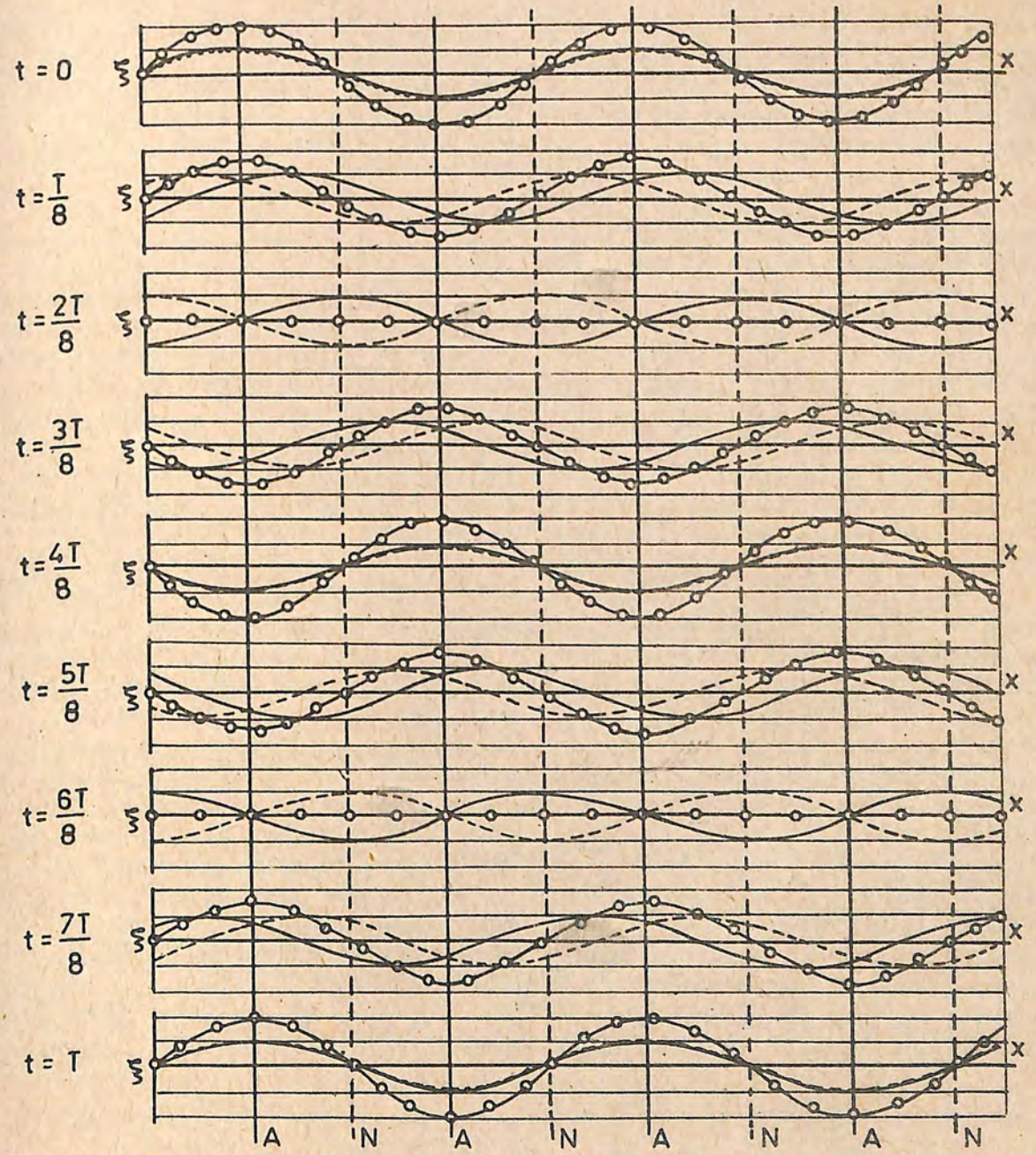


Fig. 6.7 Showing the superposition of two waves of equal λ and equal amplitudes travelling in opposite directions. 8 stages at intervals T/8 each are shown.

intervals and deduce the resultant by adding $\xi_1 + \xi_2$ in the graph at each x. In Fig. 6.6, eight steps at intervals T/8 each are shown. The full line ξ, x curve moves to the right by $\lambda/8$ in each step of time T/8. The dotted ξ, x curve moves to the left by $\lambda/8$ in each time step T/8. The resultant curves are shown beaded and are obtained by just adding $\xi_1 + \xi_2$ at each x. At $t = 0$, the curves are matching peak-to-peak state, while at $t = 2T/8$, they have peak-and bottom-relations, and at $t = 4T/8$, again, there is

peak-to-peak matching. Thus, at t = 0 and t = 4T/8, there are maximum displacements, while at t = 2T/8 and t = 6T/8, all the particles are simultaneously at zero displacement (not zero velocity).

The resultant displacement at stages t = 0, T/8, 2T/8, ... 8T/8 are shown separately in Fig. 6.8. It is to be noted that the amplitude of oscillation is zero at points like N, N, N. These are called *nodes*, or positions of zero amplitude of displacement (also of velocity). Also, the amplitude is *maximum* at points like A, A, A. These are called *antinodes*, or positions of maximum amplitudes of displacement (also of velocity). In a string, N and A are a set of points; on the water surface or in a membrane, they are a set of lines; and in space (like air), they form a set of planes—nodal planes and antinodal planes.

It will also be noted that all the particles reach their maximum displacement simultaneously. They also reach zero displacement simultaneously. In other words the phase does not vary with x, it is the same at all places and changes with time simultaneously for all positions. In that sense, there is nothing like a wave front, as in travelling (or progressive) waves.

In Fig. 6.7, the origin x = 0 is not marked. It could be at some node or at some antinode. In practice, the two oppositely travelling waves are considered as arising from reflection at some *boundary*. Such reflection can either be 'without change of sign' or 'with change of sign' (see Unit 5). In the first case, at the boundary (x = 0) the signs of ξ_1 and ξ_2 are always the same and, hence, that position has an antinode. In the second case, at the boundary the signs of ξ_1 and ξ_2 are always opposite and, hence, that position has a node.

Characteristics of Stationary Waves

In progressive waves, the amplitude is the same at all positions, but the maximum displacement is achieved by different points at different intervals of time. In contrast, in stationary waves Figs. 6.7 and 6.8 show that the amplitudes vary from place to place, being zero at the nodes and maximum at the antinodes. The distance between the two successive antinodes or two successive nodes is $\lambda/2$, and the distance between a node and its nearest antinode is $\lambda/4$. The maximum displacement is achieved by all the points at one time, zero displacement is also achieved at one time, and so on. This means that the phase does not vary with the distance x, i.e., at a given time the phase is the same at all positions.

In a stationary wave, twice during each cycle the whole medium has zero displacement, so that the potential energy is zero, and all the wave energy must be kinetic. Again, twice during each cycle the whole medium attains the maximum displacement (which is different at different distances), and at these instants the kinetic energy is zero and all the wave energy takes the poten-

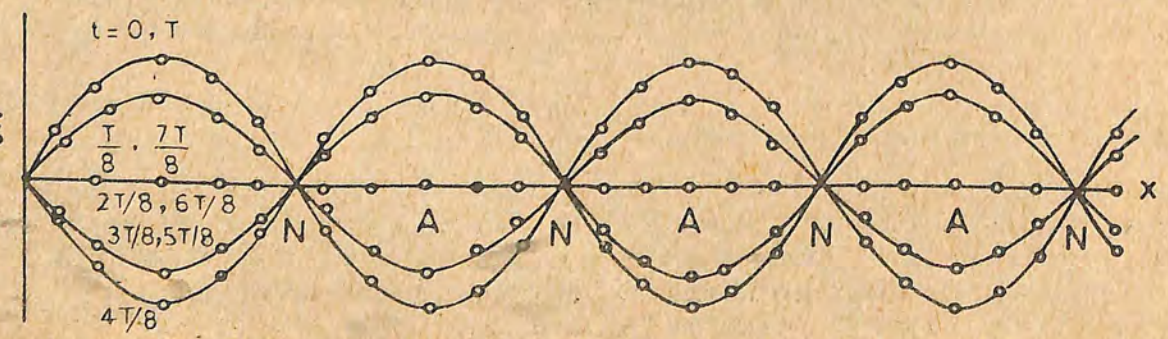


Fig. 6.8 The resultant curves of Fig. 6.7 in the stages from 0T/8 to 8T/8 (in steps of T/8).

tial form. Thus, in stationary waves, the oscillation energy alternately becomes wholly kinetic and wholly potential. In progressive waves, the energy averaged over one wave length of the medium is always half kinetic and half potential.

There is an additional point in longitudinal stationary waves. The excess pressure in longitudinal waves is proportional to the slope of the ξ, x curve (see Unit 5). In Fig. 6.8, we note that this slope is always zero at the antinodes A, A and maximum at the nodes N, N. Hence, the points A, A have no variations of pressure and may be called *pressure-nodes*. Similarly, the points N, N have maximum variations of pressure and may be called *pressure-antinodes*.

6.5 Waves in Strings and Air Columns

In an infinitely long string or in totally open air, the waves travel one way or the other with a characteristic velocity. But now our interest will be in a string bounded by two bridges, or an air column bounded by the ends of a tube.

A sonometer wire is in the first category, and its discussion is useful for the understanding of all stringed musical instruments—sitar, violin, ektara, etc. A resonance air column is in the second category, and its discussion is related with musical instruments like flute, pipe, etc.

Sonometer

It consists of a wire ABC pegged at the end A, passing over a pulley at B and carrying a load at the end C. Two sliding bridges D and E touch the wire and our discussion is for the vibrations of this bounded part of the wire DE. The mass per unit length of the wire m and the tension T control the velocity c of the waves on this wire, the relation being

$$c = \sqrt{T/m} \qquad \ldots (6.17)$$

But any disturbance created on the wire is reflected by the bridges D, E and, hence, only stationary vibrations can occur there, not travelling waves.

Without the need to go into the details of superposition of oppositely travelling waves and the formation of stationary waves, we note at once the fact that there must be a node at the end D and a node also at the end E. This is because the wire is in contact with the bridges. Now in stationary waves, the simplest condition satisfying these condition

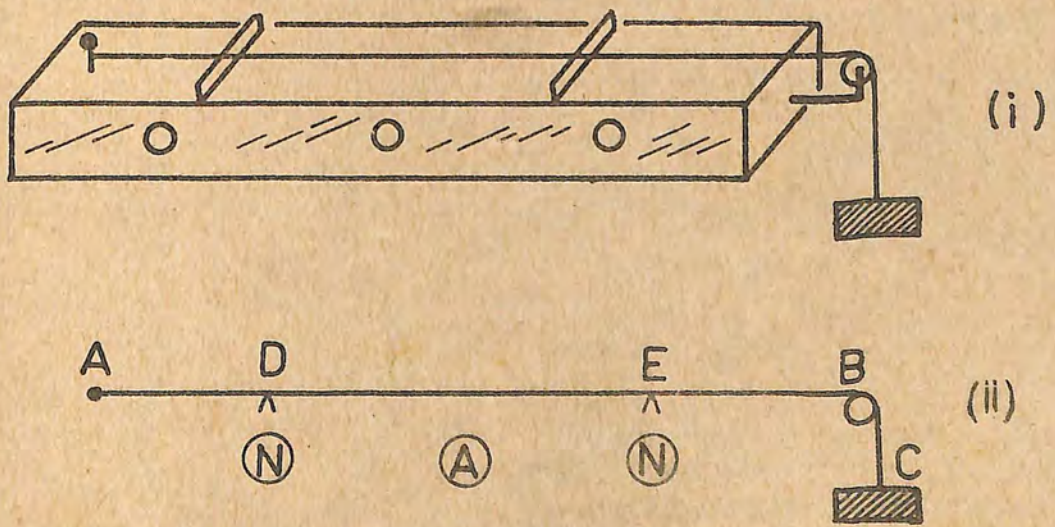


Fig. 6.9 (i) A sonometer, (ii) its schematic presentation.

is to have one antinode in between these nodes. If the length of the wire between DE is L, then N to the next N is $\lambda/2$, so that we get

$$\frac{\lambda}{2} = L \quad \text{or} \quad \lambda = 2L$$

The frequency is then given by using the equation (6.17):

$$\nu = c/\lambda = \frac{1}{2L}\sqrt{\frac{T}{m}} \qquad \ldots (6.18)$$

Thus, for a given tension and a given wire, the frequency is inversely proportional to length L. The increase of tension increases c and, hence, increases the frequency. The main point is that a stretched wire limited in length vibrates at a characteristic frequency. In any musical instrument involving chords or wires, a set of frequencies is arranged by controlling the lengths and tensions.

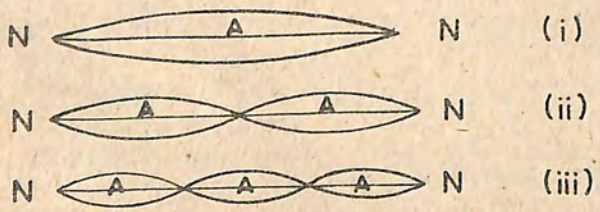


Fig. 6.10 Fundamental (i), higher harmonics (ii) and (iii) in sonometer oscillations.

Under the conditions that the ends are nodes, we could also have two or three or more antinodes in between. Fig. 6.10 shows some cases. If there are p intermediate antinodes, the wave length becomes 2L/p, and the frequency becomes

$$\nu_P = \frac{c}{\lambda} = \frac{p}{2L}\sqrt{\frac{T}{m}} = \nu p \qquad \ldots (6.19)$$

The frequencies possible in a sonometer wire are, thus, ν, 2ν, 3ν, ... The first one is called *fundamental*, and the others are called *harmonics*—second harmonic, third harmonic, etc.

EXAMPLE 6.5

A sitar wire is under tension 40 N, and the length between the bridges is 70 cm. A 5-metre sample of that wire has mass 1.0 g. Deduce (i) the speed of the transverse waves on the wire, (ii) the frequency of the fundamental, and (iii) the frequencies of the first two harmonics.

Solution

$$m = \frac{.001\ \text{kg}}{5\ \text{metre}} = 0.0002\ \text{kg/metre}.$$

$$c = \sqrt{\frac{T}{m}} = \sqrt{\frac{40}{0.0002}} = 4.5 \times 10^2$$

$$\nu = \frac{c}{2L} = \frac{4.5 \times 10^2}{2 \times 0.70} = 320\ \text{s}^{-1}$$

$\nu_P = p\nu = 640$ and $960\ \text{s}^{-1}$, for $p = 2$ and 3.

Resonance Air Column

It consists of a vertical glass tube PQ (Fig. 6.11) connected to a water reservoir R

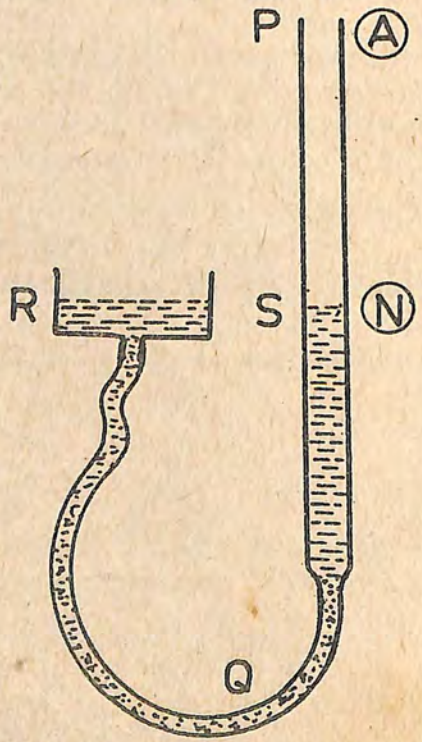


Fig. 6.11 The resonance air column.

by a rubber tube. The water level S in the tube can be altered by changing the height of R. Our interest is in the air column PS, the lower end of which is bounded by water, and the upper end bounded by free atmosphere.

The boundary conditions are: a node at the end S, because water is a rigid boundary relative to the air in the tube; an antinode at the end P, because open atmosphere is a free or loose boundary relative to the tube air. So we mark N (for the node) at S and A (for the antinode) at the end P. In the simplest conditions, no other node or antinode needs to come between the length PS. Therefore, if the length of the column is L, we get

$$\frac{\lambda}{4} = L, \quad \lambda = 4L \qquad \ldots (6.20)$$

The velocity of the sound in a gas depends on its volume elasticity E and density ρ. It is given by Newton's formula

$$c = \sqrt{\frac{E}{\rho}} \qquad \ldots (6.21)$$

Laplace suggested that adiabatic elasticity should be used for E. It is equal to γ times the pressure P, so that

$$c = \sqrt{\frac{\gamma P}{\rho}} \qquad \ldots (6.22)$$

Using this, the frequency of the fundamental vibration in the resonance column is

$$\boxed{\nu = \frac{c}{\lambda} = \frac{c}{4L}\sqrt{\frac{\gamma P}{\rho}}} \qquad \ldots (6.23)$$

Unlike the sonometer case, here c is not controllable. So, we only record that frequency is inversely proportional to length*. Among common experiences, one notes that as water is poured into any jar the pitch of the sound goes on increasing as the air column in the jar becomes smaller. Jaltarang, a musical instrument, is made of a set of cups filled with water up to different levels so as to give a musical set of frequencies.

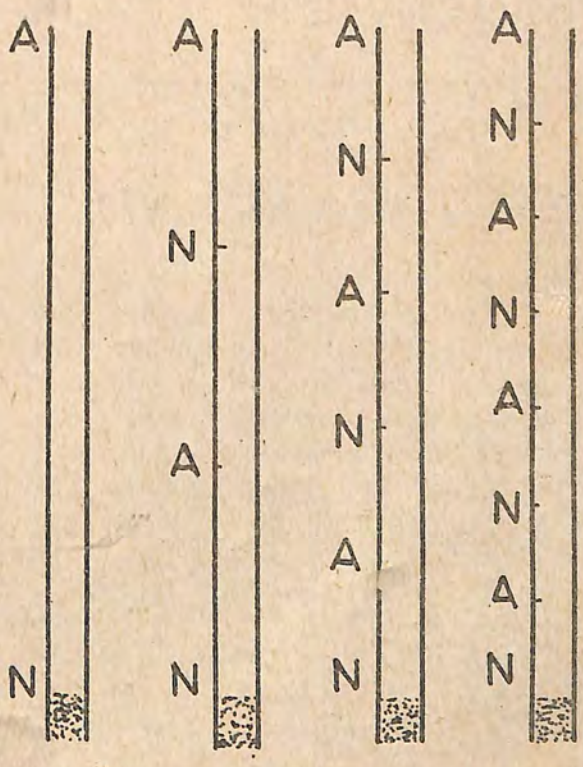


Fig. 6.12 The harmonics in a resonance air column.

Fig. 6.12 shows the higher harmonics which may occur in a resonance column. To satisfy the end-conditions, the column may be split into 1 or 3 or 5 or ... (2p + 1) ... parts, each part depicting $\lambda/4$. Thus, the general frequency is

$$\nu_p = \frac{c}{4L}(2p + 1)$$
$$= (2p + 1)\,\nu \qquad \ldots (6.24)$$

* As a matter of fact, the antinode is not just at the upper edge but a little height α above, where α is called the end correction.

i.e., apart from the fundamental frequency ν, we get the third, fifth, ... (all odd) harmonics.

Flute

A flute forms another example of a bounded air column. For simplicity, we shall first treat all the side holes as closed (Fig. 6.13).

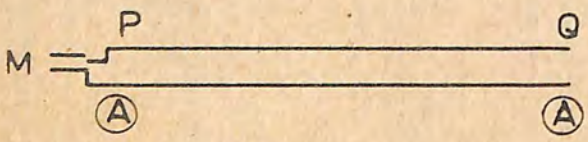


Fig. 6.13 A flute, with side holes considered closed.

The flute is blown at the end M. The opening at P and the other end Q become antinodes, because the air column connects there with the open atmosphere. Between these antinodes, there will at least be one node (the case of fundamental vibrations), but there could be 2, 3 ... p ... nodes (the cases of harmonics). The side holes are of great importance, because their positions decide which of the harmonics are to become prominent. It will be noted that a sitar and a flute are of the same character as regards the harmonics—only the nodes in a sitar exchange for the antinodes in a flute.

Vibrations of Rods : Tuning Fork

In Fig. 6.14 we show a rod rigidly clamped at the mid-point. If it is rubbed along its length, it starts vibrating longitudinally. Since both ends are 'free' and the mid-point is 'rigid' we have $\lambda = 2L$. and $\nu = c/2L$, where c is the speed of longitudinal waves in the rod and is given by $\sqrt{E/m}$, in which E is Young's modulus and m the mass per unit length of the rod.

The same rod could execute transverse vibrations also, with $\lambda = 2L$ and $\nu = c/2L$, but in that case the speed c depends in a complicated way on the cross-section of the rod, apart from elasticity. A notable case of

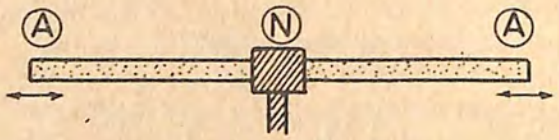


Fig. 6.14 For a rod clamped rigidly at the middle, $\nu = c/2L$

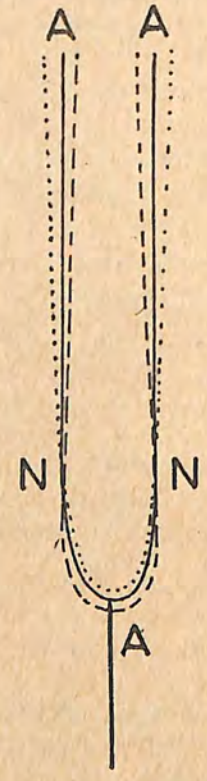


Fig. 6.15 For a tuning fork, not only the ends are antinodes, but the stem-point is also an antinode.

transverse vibrations occurs when a rod of rectangular cross-section is bent in U=form and a stem provided to it at the middle (Fig. 6.15). This is the tuning fork and detailed analysis shows that the stem-point is an antinode. Since the free ends are also antinodes, there are two nodes in—between. In this case there is no simple formula for the frequency ν. But the important fact is that whereas higher harmonics may occur quite appreciably in the vibrations of straight rods, the tuning fork gives quite a pure note, i.e., the harmonics of higher frequency are absent.

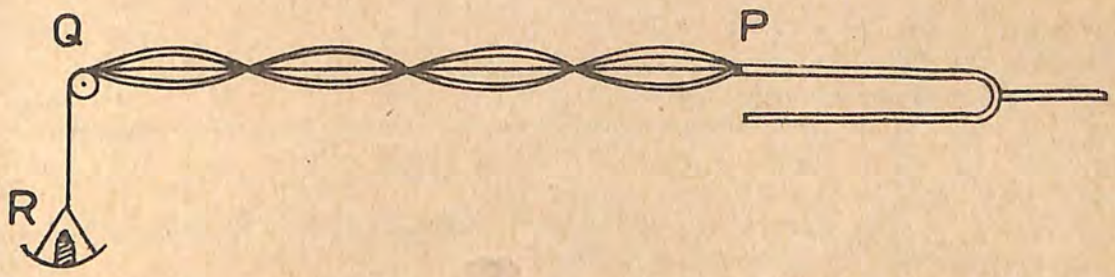


Fig. 6.16 Melde's experiment.

Resonance Experiments

A tuning fork has a fixed frequency ν_o, while a sonometer or a resonance air column (Fig. 6.11) has its frequency ν variable with L. If the vibrating tuning fork is set to transfer energy to the sonometer or the air column, the latter does not take up appreciable oscillations if ν is far from ν_o. If ν comes near ν_o, then the system takes up oscillations in some strength. If $\nu = \nu_o$, then the oscillations from the tuning fork are accepted best by the sonometer wire or the air column. The name resonance is given to this phenomenon. We may call the tuning fork a driver and the sonometer or air column a 'driven' system. Strong oscillations are generated in the driven when its frequency matches the frequency of the driver. We then say that the driver and driven are in resonance. It is like our radio circuit being set in resonance with (say) Delhi A station.

The resonance between the tuning fork and a sonometer wire leads us to write,

$$\nu_o = \nu = \frac{c}{2L} \qquad \ldots (6.25a)$$

where ν_o is the frequency of the tuning fork, and ν the frequency of the sonometer wire. Since c can be deduced from the tension and the mass per unit length of the wire, this step enables us to obtain ν_o experimentally.

Similarly, for resonance between the tuning fork and an air column (Fig. 6.11),

$$\nu_o = \nu = \frac{c}{4L} \qquad \ldots (6.25b)$$

In this case, from the known ν_o one can get the velocity of sound in air.

Melde's Experiment: It is a good experiment for seeing the stationary waves and the occurrence of several harmonics in a given string. Fig. 6.16 shows the arrangement. PQR is a soft chord. Its end P is attached to an electrically maintained vibrator (a tuning fork shown in the figure for simplicity). The thread passes over a pulley Q and carries a variable load at R. In an experiment, one may vary the tension T (which is given by Mg, if M is the mass suspended at R) and keep the length PQ = L constant. In general, the chord does not show strong vibrations. But at particular values of the tension it shows up vibrations of large amplitude. The chord may show up 1, 2, 3... p loops (Fig. 6.16 shows the case of 4 loops). The amplitude may become as high as ~ 1 cm and, hence, it is a good experiment for demonstration.

Kundt's Tube: The velocity of sound in a gas other than air is measured using a Kundt's tube (Fig. 6.12). ABC is a rod (like that in Fig. 6.14) clamped rigidly at the middle, and it may be vibrated longitudinally. PQ is a long glass tube containing the gas. C is a plate attached to the rod and losely fitting the tube, whereas R is an adjustable piston at the other end. Dry lycopodium powder is spread along the tube.

The rod is now vibrated and piston R is adjusted till we find that vibrations of the

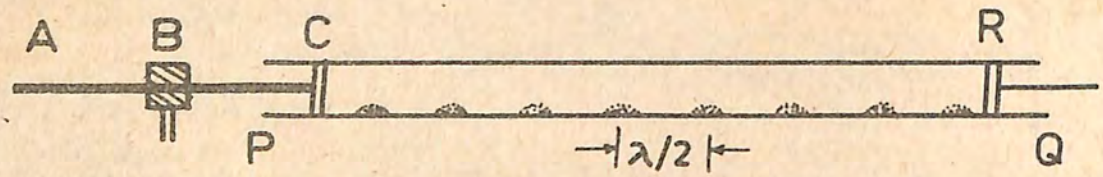


Fig. 6.17 Kundt's tube.

rod make the lycopodiam particles dance appreciably. This shows resonance between the vibrations of the rod and of the gas in the tube. Repeated excitations of the rod then make the lycopodium powder collect into heaps as shown, and their separation gives $\lambda/2$. If the experiment is done once with air and then with the gas in the tube, we have

$$\nu_{rod} = \frac{c\,(\text{in air})}{\lambda\,(\text{in air})} = \frac{c\,(\text{in gas})}{\lambda\,(\text{in gas})}$$

$$\frac{c\,(\text{in gas})}{c\,(\text{in air})} = \frac{\lambda\,(\text{in gas})}{\lambda\,(\text{in air})} \qquad \ldots (6.26)$$

EXAMPLE 6.6

In a sonometer chord of length 24.7 cm, resonance occurs with a tuning fork of the frequency 256 Hz. Deduce the velocity of waves over the chord. At what length will resonance be observed with a fork of $\nu = 453$ Hz?

Solution

We assume that resonance is occurring with the fundamental vibration. Then

$$\nu = \frac{c}{2L},$$

$$c = 2\nu L$$

$$= 2 \times 256\ s^{-1} \times 24.7\ cm$$

$$= 126\ m/s.$$

For the second case, let L be the length at resonance. Then, since c is the same for the given chord under the given tension,

$$\frac{L_2}{L_1} = \frac{\nu_1}{\nu_2},$$

$$L_2 = 24.7 \times \frac{256}{453}\ cm$$

$$= 14.0\ cm$$

EXAMPLE 6.7

A resonance air column shows resonance with a tuning fork of $\nu = 256$ Hz. at column length 33.4 cm and 101.8 cm. Deduce (i) the end-correction, and (ii) the speed of sound in air.

Solution

If α is the end-correction, and L_1, L_2 the lengths at the successive resonance points with a driver of frequency ν.

$$c = 4\nu(L_1 + \alpha) = \frac{4\nu}{3}(L_2 + \alpha)$$

Now to get α, we note that

$$L_1 + \alpha = \frac{L_2 + \alpha}{3}$$

$$\therefore\ \alpha = \frac{L_2 - 3L_1}{2}$$

$$= \frac{101.8 - 3\,(33.4)}{2} = 0.8\ cm$$

Next,

$$c = 4 \times 256\,(33.4 + 0.8)\ \text{cm/sec.}$$
$$= 350\ \text{ms}^{-1}$$

6.6 Acoustic Considerations in Everyday Life

Some topics of general interest concerning sound will be very briefly discussed here. Light and sound are two means of perception from a distance. While for light, all human beings depend on external sources; for sound, each person has his own vocal chord, with almost an infinity of depth and variants. A study of the vocal chord is done under physiology, but a study of its product—the voice—is acoustics. We shall refer to some broad features only.

Frequency of sound waves which the average human ear can hear ranges from 20 Hz to 20,000 Hz. One does not hear oscillations outside this range, but within this range the ear is quite capable of perceiving the difference of frequency. What is called high pitch corresponds to high frequency. The study of the human ear and its frequency perception is another interesting area. The outer ear collects the oscillations from a larger area and passes them to the eardrum of much smaller area; then, there are resonators of some kind which respond to different frequencies and carry the message to the brain.

In this context, it is interesting that the range of frequencies for hearing is not the same for various insects and animals as for human beings. Bats, for example, produce and hear sounds of a much higher frequency than 20,000 Hz. In fact, they use sound waves as their 'eyes' for sensing the surroundings. Dogs are known to hear frequencies $> 20{,}000$ Hz, so that shikaris use a special whistle which the human ear cannot hear but the Shikar dogs do.

Another characteristic of sound is its intensity—the energy flow per unit area per second. The human ear can respond to a very wide range of intensities—a ratio of over 10^{12} from the weakest sound to that which starts giving pain. To realize this, one may notice that at one end the sound of a water drop falling from a small height on a tin-sheet can be heard from a distance of several metres, and at the other end the sound of a heavy hammer dropped from a large height by a blacksmith is tolerated by him though he is so near.

We have studied about the interference of waves, which depends on the phase difference between two waves. The everyday use we make of it is through the fact that we have two ears. Depending on the direction of the source relative to our two ears, there will be a phase difference between the waves reaching the two ears. Conversely, from the sensing of this phase difference, one senses the direction of the source of the sound. In fact, one often turns his head a little to change the phase difference so as to reach a better conclusion.

We had talked in the last unit about the reflection of sound. Echoes are probably the best examples—echoes from the distant hills and echoes from the distant buildings. But more significant is the fact that by measuring the time-gap between the production of a sound and the reception of an echo one measures the distance of the reflector. Depths in lakes and oceans are found thus, and ships locate the under-sea rocks by this method*. In nature, the bat (which has no eyes) uses this sensing device all the time to see its surroundings.

A distinction between music and noise may also be of interest. Musical sounds are periodic, while an aperiodic sound is noise. A pure sinusoidal wave is musical. If two waves of a simple frequency ratio (like 1 : 2,

* Actually, very high frequency waves (ultrasonics) are used here.

2: 3, 3: 5, etc.) combine, the resultant is still periodic (see Fig. 6.18a), and hence musical. But if two sinusoidal waves of an odd ratio (like 1537: 1385) combine, their resultant is aperiodic and, therefore, becomes 'noise'. In the human voice, the vowel sounds are 'musical', while the consonants form 'noise'. A flute is musical because although we just blow it by the mouth, it resonates to some frequency ν and its harmonics; so does every musical instrument—the energy supply is not periodic, but the system gives out periodic oscillations of chosen frequencies. Two unmatched musical instruments may generate 'noise'. The worst noise is that which contains a mix of all frequencies—that is what we have in the rattling of a factory.

Since sound recording and reproduction are so widely used, it is desirable to emphasize some points of special interest. The human voice produces oscillations with combinations of widely varying frequencies and wide proportions of amplitudes. A good device (radio, telephone, amplifier, gramophone, etc.) is that which reproduces all these details exactly in the same proportions. This property is called *fidelity* (which means faithfulness). If any one out of the components (like the microphone, the amplifier and the loudspeaker) has poor fidelity, then the reproduced sound will not match the original. In a radio, often a *tone control* is provided, so that one may adjust the proportionate contribution from low, mid and high frequency components in the sound.

We shall now pay some attention to the acoustics of a building. One property we need is that there should not be much penetration of sound between rooms. In hotels and studios of radio stations, this is a special requirement. For this, the walls must be covered with sound absorbant materials and the doors must have heavy curtains on them. Just the opposite is the example of 'whisper-

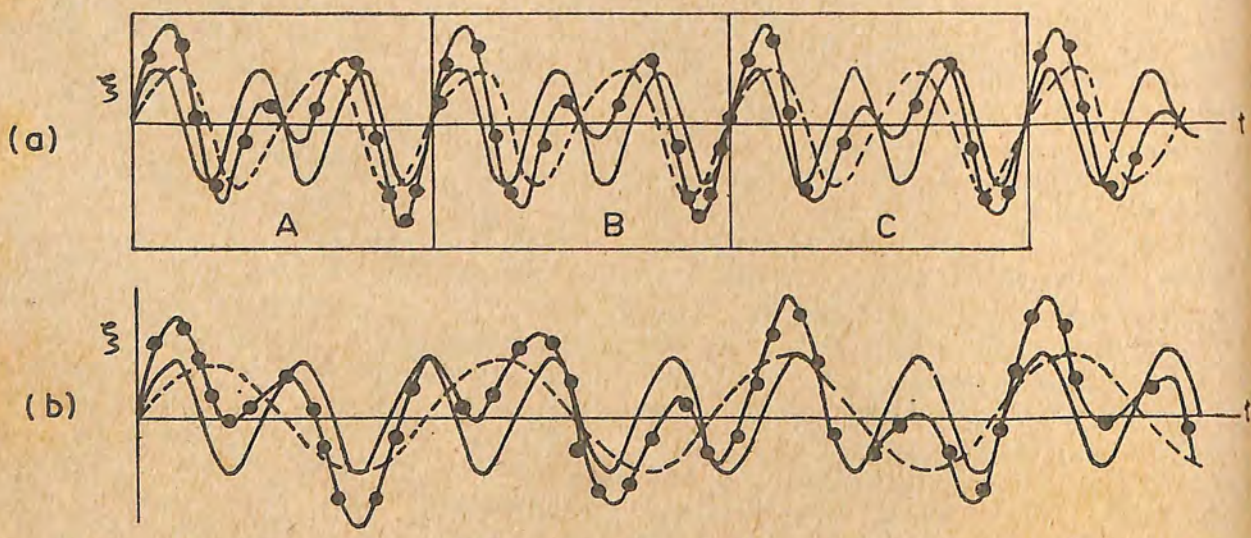


Fig. 6.18 (a) Combination of waves in a simple frequency ratio (2:3 here) produces a periodic oscillation blocks A, B, C are identical; (b) combination of waves in an odd frequency ratio (10:23 here) produces a disturbance in which the repetition pattern does not appear in small enough time.

* Note that 'echo' is the return of the sound of a syllable after the original sound is dead, while 'reverberation' is the continuation of the sound of a syllable for some time.

ing galleries', in which the walls are so hard (hence, non-absorbant) that even whispers can be heard at diametrically opposite ends of a large gallery around a dome.

In large auditoriums, echoes create a problem. The average interval between two syllables spoken by a person is around 0.2 second. If the reflected sound of any syllable comes after the next syllable directly reaching a listener, then there is confusion. If the reflection comes from a wall or ceiling distant d, the time taken is 2d/c, and what we want is that if 2d/c is greater than 0.2 second, then that wall or ceiling should be made absorbant (i.e., non-reflecting).

Another problem is reverbération. A sound once produced repeatedly reflects in the room or hall and dies out only gradually. If there are lots of doors and windows or absorbing walls or heavy curtains or soft furnishings, then the sound is absorbed fast, so that it dies out quickly. But in an unfurnished hall with not much audience and open window area, the sound may persist for long. The time taken by the sound intensity to fall by a factor 10^{-6} is called the *reverberation time* t_R of that hall*. It may be as large as 20–30 seconds for the central dome of the Taj Mahal and as small as 0.2 second for a rich hotel room. Now, if t_R is very small, each syllable can be heard distinctly and separately, while if t_R is too large, there will be mixing up of several syllables, leading to confusion in hearing. By experience, it is found that t_R about 1.0 second is quite good for hearing. That is an average. For a conference, smaller t_R is desirable. For a musical concert, a little larger t_R is quite pleasant. For a given hall if t_R is too large, we should introduce a heavy carpet, a false ceiling, curtains, absorbing coverage on walls, etc. It

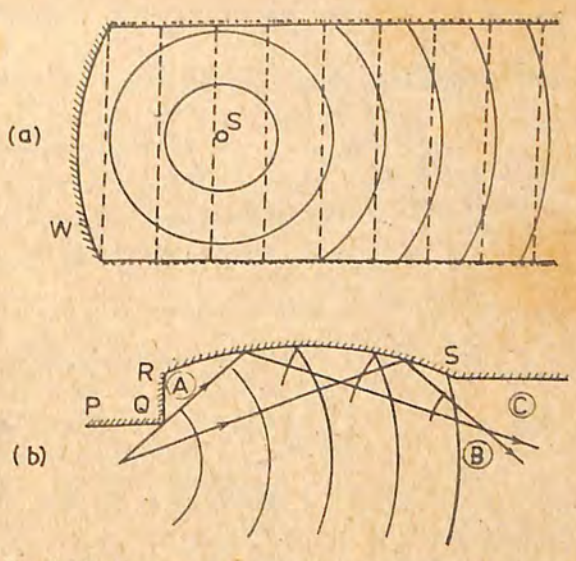


Fig. 6.19 (a) A parabolic wall at the speaker's back helps in even distribution of sound; (b) Focusing by other walls leads to uneven distribution.

is also notable that the audience itself reduces t_R because of the sound-absorbing area—particularly by women in saris and shawls.

For a lecture hall, a special problem is uneven distribution of sound. One method used to get even distribution is to have the wall W (Fig. 6.19a) in front of the audience parabolic, with the speaker at its focus. The reflected sound from this wall then goes equally to every one (Fig. 6.19a). But if other walls are curved, they could contribute to improper focusing. In Fig. 6.19b, for example, the wall RS focuses the reflected sound in the region B at the cost of the regions like C. There are also certain 'dead centres' where sound neither reaches directly nor by reflection. In Fig. 6.19b, the region A is affected by the 'shadow' from the projection Q. But on the side far from the speaker, the reflected sound is so significant that its absence may make a region 'dead'.

EXERCISES

6.1 State the principle of superposition. Is it valid in the case of light? Draw two ξ, t sinusoidal curves of periods T and T/2, taking the amplitude of the latter one-half that of the former. Deduce the resultant ξ, t curves over the total time 3T.

6.2 (a) Among two interfering sources, let S_1 be ahead of the phase by 70° relative to S_2. If an observation point P is such that $PS_2 - PS_1 = 1.5\,\lambda$, deduce the phase difference between the waves from S_1 and S_2 reaching P.

$[\phi_1 - \phi_2 = (3 + 7/18)\,\pi]$

(b) Two vertical antenna have separation $= \lambda/4$, and their oscillations have the phase difference $\pi/2$. Discuss the distribution of intensity in the horizontal plane.

6.3 (a) **In an experiment with Quinckes tube, one passed from one maximum to the next** maximum sound when the U-tube was slided by 30 cm outwards. Deduce the wave length and frequency of the sound wave, given $c = 350\ ms^{-1}$

(60 cm, $5.83 \times 10^3\ s^{-1}$)

(b) Two sources of an equal frequency produce interference. Now one source is slightly loaded so that it gets behind the phase at the rate of $\pi/10$ radian per second. What will be observed at a given point in space as time passes?

6.4 (a) Prove that if a_1 and a_2 are the amplitudes of two interfering waves and ϕ their phase difference, the amplitude a of the resultant vibration is given by

$$a^2 = a_1^2 + a_2^2 + 2\,a_1\,a_2 \cos\phi$$

(b) Two interfering sources have an intensity ratio 16 : 1. Deduce their amplitude ratio and the ratio of intensity between the maxima and minima in interference.

(4 : 1 ; 25 : 9)

6.5 Define diffraction. Why is it found so common in sound, while it is not common in light? In the ripple tank experiment (Fig. 6.4), what would be observed if (i) an obstacle $< \lambda$ in width is used, and (ii) an obstacle $\sim 2\,\lambda$ in width is used, instead of the slits. (If necessary, do the experiment.)

6.6 (a) A sitar wire and a tabla, when sounded together, give 4 beats per second. What do we conclude from this? As the tabla membrane is tightened, the beat rate may decrease or increase. Explain.

(b) An observer moving towards a wall at 20 m/s hears a sound from a source at some distance behind him directly as well as after its reflection from the wall. Deduce the beat frequency between these two sounds, taking the true frequency at 580 Hz and $c = 340$ m/s.

(10 beats/s)

6.7 (a) The fact that we have two ears enables us to sense the direction from which a sound is coming. This depends on the interference principle. Explain.

(b) Sending a sound wave and receiving a scattered wave from some object is a method

of detecting the object. The distance d is estimated from the time-gap t. Explain how the direction may be estimated. Even more, the speed of the scattered object may be determined by the beat frequency between the original wave and the scattered wave (cf. Doppler effect). Explain.

6.8 Examine the following equations, each representing a wave:

$$\xi_1 = a\cos 2\pi\left(\frac{t}{T}+\frac{x}{\lambda}\right)$$

$$\xi_2 = a\sin 2\pi\left(\frac{t}{T}-\frac{x}{\lambda}\right)$$

$$\xi_3 = a\cos 2\pi\left(\frac{t}{T}+\frac{x+p}{\lambda}+\phi\right)$$

$$\xi_4 = a\cos 2\pi\left(\frac{t}{T+\alpha}+\frac{x}{\lambda}\right), \alpha \ll T.$$

List whether interference, beats or stationary waves will arise from the superposition of (i) 1 and 2, (ii) 1 and 3, (iii) 1 and 4, (iv) 2 and 3, (v) 2 and 4, (vi) 3 and 4.

6.9 Define nodes (N) and antinodes (A) in stationary waves. Do the pressure nodes and pressure antinodes coincide with N and A, respectively? In terms of λ, what is the separation of a node from its nearest antinode? What is the phase difference between points separated by $\lambda/10$ in stationary waves?

6.10 (a) Compute the fundamental frequency of a sonometer wire of the length 20.0 cm $T = 20\,N, m = 5.2 \times 10^{-3}$ kg/m.

(150 s^{-1})

(b) A resonance air column resonates with a tuning fork of the frequency 512 Hz at the length 17.4 cm. Neglecting the end-correction, deduce the speed of sound in air. Is your answer unique for the given data?

(356 m/s)

6.11 (a) A pipe of the length 2 L, open at both ends, has the same frequency as another pipe of the length L closed at one end. Prove this. Also, state if the total sounds will be identical for the two pipes.

(b) Discuss the possible harmonics in the longitudinal vibrations of a rod clamped at the middle. Compare it with the cases of (i) a sonometer, (ii) a pipe open at both ends.

6.12 In a tuning fork, the stem point is not a node. This result follows from the fact that the centre of mass of an isolated system should not oscillate. Try how the logic works.

6.13 (a) **In the equation (6.15), v and v_0 refer to two different quantities. Comment.**

(b) In Melde's experiment (Fig. 6.16), the end P attached to the vibrator is nearly a node, not an antinode, although the vibration energy is fed from there. Explain.

एकलव्य
अध्ययन केन्द्र मालाखेड़ी
औरंगाबाद फोन- 277267

(c) In Melde's experiment (Fig. 6.16), should λ be obtained by measuring the distance PQ, or the distance between any two intermediate nodes?

6.14 Two main characteristics of a sound amplifying system are (i) high response, (ii) fidelity. Discuss the meaning and significance of 'fidelity'. What is the function of a 'tone control' in a radio receiver?

6.15 (a) Discuss the defects in a hall from the acoustic point of view.

(b) Define the reverberation period t_R. Why should it be neither too high nor too low for good hearing? Consider a musical concert, a lecture, and a conference. In which case would you prefer a somewhat larger t_R and in which case smaller? Give reasons.

UNIT 7

Optics

In this Unit, we shall pay special attention to the experiments which establish that light shows wave-like character, i.e., it shows polarisation, interference and diffraction. We have already discussed these in general under 'Wave Motion'; but in the case of light, some special details are needed, which we shall discuss here. Besides, we shall also describe how the light from a given source may be examined for its wave-lengthwise distribution. This distribution is very different for the light from the sun, from a mercury discharge tube, and from a Bunsen burner flame. The study leads to our understanding of the nature of the light emission process—which really leads to atomic and molecular physics.

7.1 Nature of Light

From our everyday experiences about shadows, we get the law of rectilinear propagation of light. Newton proposed that light may consist of 'some particles' (called corpuscles) shot out by the source and travelling without being affected by earth's gravitation. This theory is named as *corpuscular theory*.

From our present day knowledge, one serious flaw in this theory is seen from the measurement of the velocity of light. It is found that this velocity does not depend on the source at all, it has a definite value for each medium (for a given colour of light). Could it be that all sources somehow manage to throw the corpuscles with the same speed? The more reasonable assumption would be that the source just releases some disturbance, and the velocity of transmission is a property of the medium, not of the source.

The second serious flaw in the corpuscular theory is seen from the experiment on polarisation described in the Unit on Wave Motion. The experiment leads to the transverse wave character of light. Whether the transverse 'property' is a displacement of 'ether' (as imagined in earlier theories) or an electric or magnetic field vector (as taken in the electromagnetic theory) will be settled by other experiments, which we shall now take up.

Waves show interference, which is *ordinarily* not seen with light. Waves also show bending round corners (diffraction), which is *ordinarily* not seen with light. Careful experiments now show that both these phenomena do occur very widely and in exactly the same way as in sound waves—only the wavelength of light is very small relative to our familiar dimensions and, hence, these phenomena do not become obvious in our daily experience.

7.2 Interference of Light

Among many experiments showing interference of light, the oldest and the simplest is Young's double slit experiment. Fig. 7.1 shows the arrangement. A is a source of light of a single colour (monochromatic), e.g., a sodium vapour lamp. S is a slit $\sim$ 1 mm wide, S_1 and S_2 are two slits at separation $\sim$ 2 mm, the width of each slit being $\sim$ 0.3 mm. OP is the plane of observation, where the light from both S_1 and S_2 reaches. The distance SS_1 is $\sim$ 10 cm, S_1O is $\sim$ 2 metre.

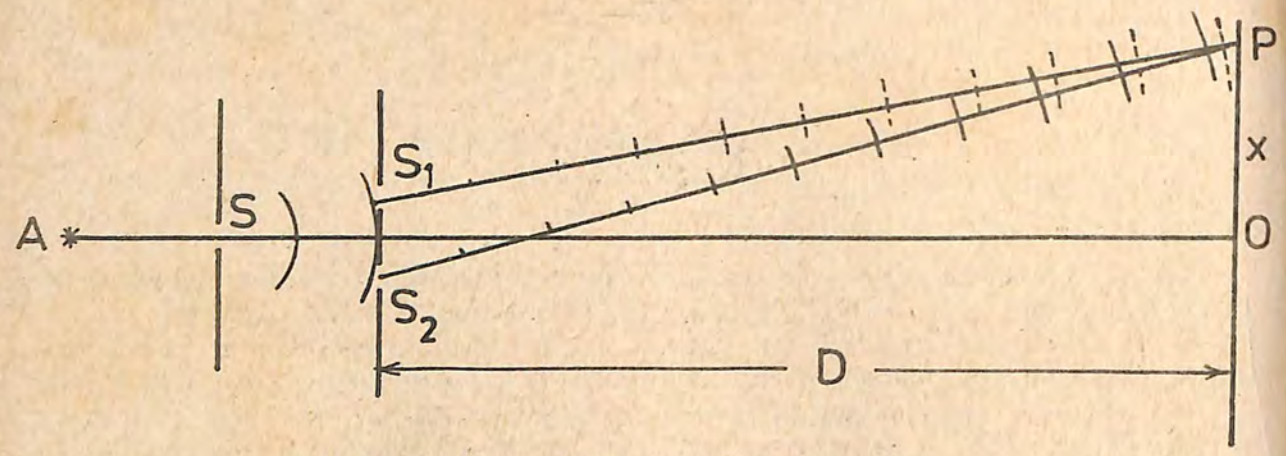


Fig. 7.1 Young's double-slit interference experiment.

The observations in the plane OP are made by an eye-piece, and one finds the following facts:

(i) If either S_1 alone or S_2 alone is open, the intensity of light is uniform on any plane like OP.

(ii) If both S_1 and S_2 are open, the intensity alternately increases and decreases as one moves along OP, we say bright and dark fringes are seen (Fig. 7.2).

(iii) The fringe width (ω in Fig. 7.2) is proportional to the distance D and inversely proportional to the slit separation $S_1 S_2 = d$.

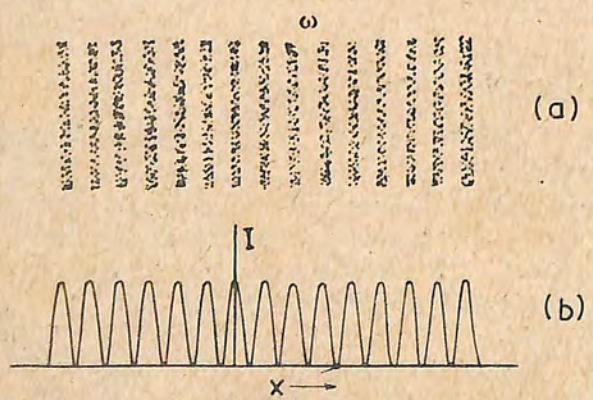


Fig. 7.2 Fringes in Young's experiment: (a) A view of the fringes; (b) representation of the intensity variation along the fringes.

To explain the observations, let us assume that light consists of some wave-motion with the wavelength λ. In that case, wavefronts from S grow out with time, disturb both S_1 and S_2, from where fresh wavefronts grow, and, finally, in the space on the right-hand side the disturbances from S_1 and S_2 would interfere, as seen in the Unit on Superposition of Waves. If we consider a point P in the observation plane, the nature of interference between the two waves reaching there depends on the *path difference* $S_2P - S_1P = p$. If $OP = x$, $S_1S_2 = d$, then*

$$p = S_2P - S_1P = \frac{xd}{D} \qquad \ldots .1)$$

If a_1 and a_2 are the amplitudes of the disturbance at P due to S_1 and S_2 separately, then with S_1 and S_2 together the amplitude at P will be $a_1 + a_2$ if the path difference p is a whole number of λ, and $(a_1 - a_2)$ if p is an odd number of $\lambda/2$. Now, on the observation plane OP, the maxima and minima should occur as follows:

$$\text{Maxima:} \quad \frac{x_n d}{D} = n\lambda \qquad \ldots (7.2a)$$

$$\text{Minima:} \quad \frac{x'_n d}{D} = \left(n + \frac{1}{2}\right)\lambda \qquad \ldots (7.2b)$$

* The deduction is given in the Unit on Superposition of waves.

The separation of the successive maxima gives the fringe width ω. Thus,

$$\omega = x_{n+1} - x_n = \frac{D}{d}(n + 1 - n)\lambda$$

$$\text{or } \omega = \frac{D}{d}\lambda \qquad \ldots (7.3)$$

We had assumed that light had wave character and got the conclusion that there should be bright and dark positions in the observation plane. Well, that is exactly what is observed experimentally. So, the assumption is satisfactory. We also note the quantitative agreement between the experimental results and the equation (7.3) that $\omega \propto D$ and $\omega \propto 1/d$. Therefore, the measurements of ω, d and D give us

$$\lambda = \frac{\omega d}{D} \qquad \ldots (7.3a)$$

It is notable that the interference experiments give us no information whether the waves are longitudinal or transverse.

It is customary to express the wavelength of the visible light in Angstron unit (A°), which is 10^{-10} m. Thus, for yellow light $\lambda = 6 \times 10^3$ A°. In the visible light the colours have the wavelength – violet ~ 4000 A°, green ~ 5600 A°, yellow ~ 5900 A°, red ~ 7500 A°. But the light of $\lambda < 4000$ A°, which is not visible to the eye is called ultraviolet; and the light of $\lambda > 7500$ A°, which again is invisible, is called infra-red. Far lower in wavelength are X-rays (a few A°) and γ-rays, (10^{-4} A°), far higher in wavelength are microwaves (~ 10^8 A°) and radio-waves (~ 10^{12} A°), i.e., 100 m. All these have essentially the same character, excepting the difference in the wavelength. One can perform interference experiments much more conveniently with microwaves because the λ is a few cm. The point is that the interference is shown by all these waves of light.

EXAMPLE 7.1

In Young's experiment (Fig. 7.1) let the $SS_1 - SS_2 = 4.75\ \lambda$. Deduce the revised forms of the equations (7.2a) and (7.2b).

Solution

The path difference p for the waves reaching P should be redefined as $(SS_2 + S_2P) - (SS_1 + S_1P)$. That gives

$$p = (SS_2 - SS_1) + (S_2P - S_1P)$$

The first part is now given as $-4.75\ \lambda$, and the second part is given by geometry as xd/D (as in the equation 7.1). Hence,

$$\text{for maxima } -4.75\lambda + \frac{xd}{D} = n\lambda$$

$$\text{or } \frac{xd}{D} = (n' + 0.75)\lambda$$

where $n' = n + 4$ = an integer. Similarly, for minima

$$\frac{xd}{D} = (n'' + 0.25)\lambda$$

where n'' is also an integer.

EXAMPLE 7.2

If the two slits in Young's experiment have widths in a ratio 1 : 9, deduce the ratio of intensity at the maxima and minima in the interference pattern.

Solution

The intensities due to separate slits are in proportion to the slit-widths. The amplitudes are proportional to the square of intensities. Calling them a_1 and a_2, we get

$$a_1 : a_2 = \sqrt{1} : \sqrt{9} = 1 : 3$$

The amplitude at maxima is $a_1 + a_2$ and at

minima $a_1 - a_2$. The intensities are proportional to their squares. Calling them I_{max} and I_{min}, we get

$$I_{max} : I_{min} = (a_1 + a_2)^2 : (a_1 - a_2)^2$$
$$= 4^2 : 2^2 = 4 : 1$$

7.3 Coherent Sources

One may ask whether interference can be seen if S_1 and S_2 in Young's experiment are illuminated by two separate monochromatic sources of the same wavelength, like two sodium lamps. It has been tried, and the result has always been negative. Two separate sources of light of the same frequency do not show interference.

The reason is now understood. Light is emitted not by bulk matter acting as a whole but by individual atoms. In a tuning fork, the whole arm oscillates as one unit; in a flute, all the air molecules oscillate in the phase. But in a sodium lamp, even 1 mm^3 volume at sodium vapour pressure 10^{-5} atm contains $\sim 10^{11}$ atoms of sodium, and the emission of light by them takes place independently*. Thus, in the case of light, a 'source' means billions of atoms which do not emit light in phase. Now, between two independent sources of light a constant phase difference has no meaning—because we really have two separate sets of billions of atoms. In the statement leading to the equation (7.2), we assumed that between S_1 and S_2 the phase difference is zero or integral multiple of 2π. If the phase difference is anything else, say ϕ, we can modify the equation (7.2) and still define maxima and minima. But if ϕ randomly changes, then maxima and minima also randomly change their positions and, thus, no interference can be seen.

In Young's device (Fig. 7.1), both S_1 and S_2 get light from the slit S. Therefore, whatever phase changes take place in one also occur in the other. For billions of waves reaching S_1 (from billions of atoms in the source), an identical set reaches S_2 also, the relative phase difference being the same for all—the corresponding wave-pairs. Such a pair of sources having a constant phase difference is called a *coherent pair of sources*. S_2 is said to be coherent with S_1. For observing interference, the sources involved must be coherent, i.e., they have constant phase relationship. In the case of light, two separate sources cannot be coherent.

Fresnel's Biprism Method

Fresnel produced coherent sources by the *biprism method* shown in Fig. 7.3. S is the slit illuminated by a source on the left. The light passes through the biprism ABC, whose upper and lower parts create virtual images S_1 and S_2, respectively, so that on the right side of the biprism the light, in effect, comes not from S but from S_1 and S_2. In the observation plane, the area GF receives light from both S_1 and S_2 and interference is seen in this area.

Llyod Mirror

This is another arrangement (Fig. 7.4) for getting coherent sources. S is the slit illuminated by a source on its left. A simple glass-plate blackened on the lower side acts as the 'mirror' M (it is not a silvered mirror), so that an area GH on the observation plane gets the reflected light from the virtual image S′ and also the direct light from S. In this area, interference fringes are observed, S and S′ being the coherent sources.

The nature of fringes and the expressions for the fringe-width are the same in Young's double slit, Fresnel's biprism and Llyod's

* In recent years, *lasers* have been developed in which the emission of light by the atoms is not independent (spontaneous) but is controlled (induced) by a signal (see § 7.7).

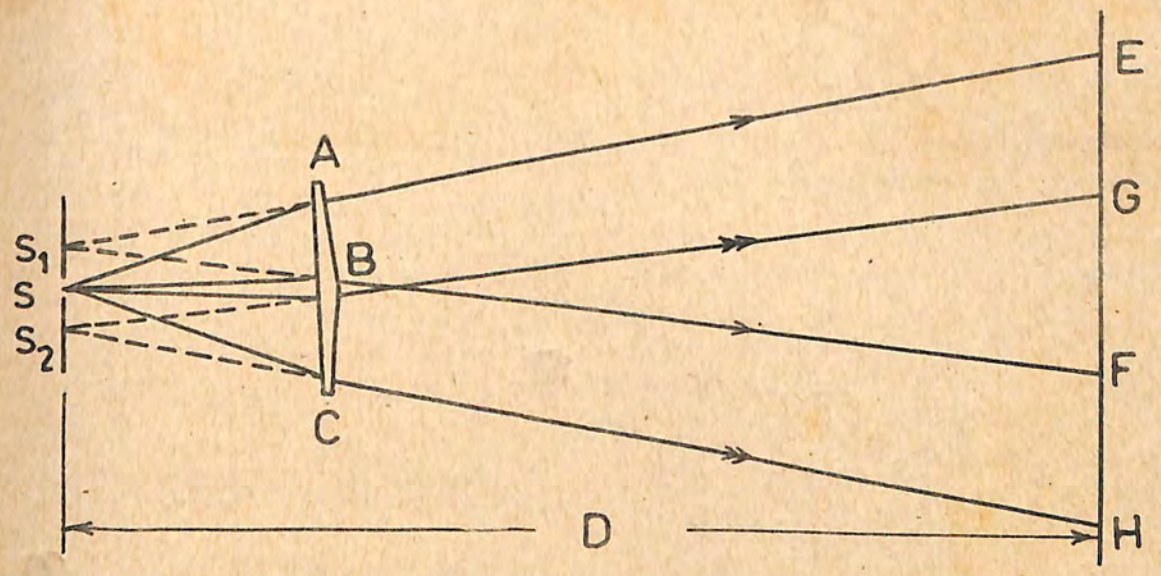


Fig. 7.3 Fresnel's biprism method of getting coherent sources and interference.

mirror* and are as shown in Fig. 7.2 and the equation (7.2). All these go under the first name alone when principles are discussed, although in laboratories the other two show fringes with much better intensity.

EXAMPLE 7.3

In a Llyod mirror interference experiment, the slit and its image have a separation 4.32 mm, and the observations made at a plane 2.00 m away show fringes of separation 0.260 mm. Deduce the wavelength of light.

Solution

The equation (7.3a) is $\lambda = \omega d/D$. From the given values here,

$$\lambda = \frac{0.0260 \times 0.432}{200} \text{ cm}$$

$$= \frac{2.60 \times 4.32}{2} \times 10^{-5} \text{ cm}$$

$$= 5.62 \times 10^{-5} \text{ cm}.$$

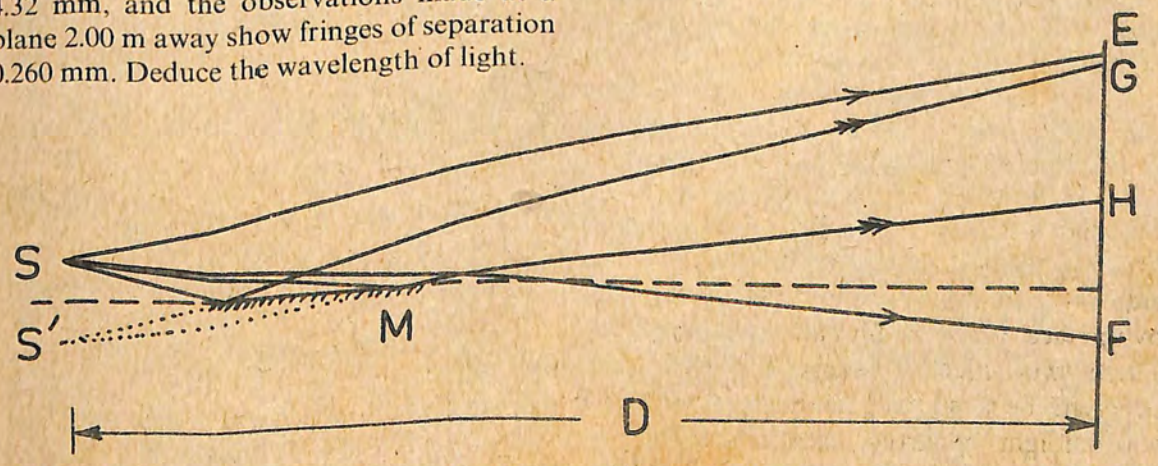


Fig. 7.4 Llyod's mirror method of getting coherent sources and interference.

* There are many other ways, too.

EXAMPLE 7.4

In a biprism experiment with light of λ = 5893 A°, we had d = 4.00 mm, D = 1.50 m. Deduce the maximum slit width so that fringes may not be obliterated.

Solution

If the slit is wide, then for each part of the slit a fringe system is formed which is shifted sideways. The fringes are obliterated when the total slit width becomes $\geq$ fringe width. From the data,

$$\omega = \frac{\lambda D}{d} = \frac{5.893 \times 10^{-5} \times 150}{0.400} \text{ cm}$$

$$= 0.22 \text{ mm}.$$

Hence, the slit must not be wider than $\sim$ 0.2 mm.

7.4 Colours of Thin Films

A soap film seen in the reflected white light shows beautiful colours. In the rainy season, the water collecting in the road-pits often gets a few drops of oil from the passing cars, and the oil film thus made on the water surface shows lovely colours. The phenomenon is fully explained by the interference of light, as we shall presently see.

In Fig. 7.5, AB and CD are the two surfaces enclosing a thin film of some medium of the refractive index μ. E is the observer's eye and by drawing backwards from E we find that for observing an area near P on the film we need a source along the line PS. However, the light from S is reflected partly by the face AB of the film and partly by CD. These two reflected beams a_1 and a_2 are coherent, because both are born of the same original light by partial reflections at the two faces of the film. In the same way, from each part of the film the eye receives two coherent beams, which, therefore, show interference.

The phase difference ϕ between the two

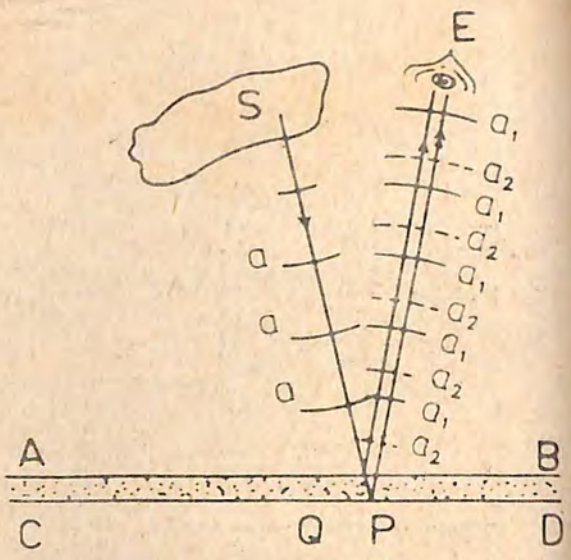


Fig. 7.5 Explaining the colours of a thin film.

reflected waves is $2\pi/\lambda$ times the path difference. We shall, for convenience, limit the discussion to light falling nearly normal to the film (in Fig. 7.5, the inclined incidence is shown for clarity). In that case if the thickness of the film at P is t, then the path difference is 2t in the medium of the film. Since the velocity of light in the medium is $1/\mu$ times that of free space, the path 2t in the medium is equivalent to $2\mu t$ in free space, so that

$$\phi = \frac{2\pi}{\lambda} \cdot 2\mu t = \frac{4\pi\mu}{\lambda} t \qquad \ldots (7.4)$$

where λ is the wavelength of the light in free space.

Now, one may at once put conditions for the maxima and the minima

$$\frac{4\pi\mu}{\lambda} t = 2n\pi \qquad \ldots (7.5a)$$

$$\frac{4\pi\mu}{\lambda} t = (2n + 1)\pi \qquad \ldots (7.5b)$$

We have not stated here which equation represents the maxima and which the minima. There is a reason for this. The equation (7.5a) should represent the maxima, but if the relative phase change π occurs between the

two reflections, in addition to that deduced from the path difference, the situation may change. In the particular case under study, the wave reflected from the upper surface will suffer a phase change of π and no such phase change occurs on the waves reflection from the lower surface. Therefore, the equation (7.5a) represents the conditions for the minima and the equation (7.5b) for the maxima. We can now put

$$\text{maxima } 2\mu t_{max} = \left(n + \frac{1}{2}\right)\lambda \quad \ldots(7.6a)$$

$$\text{minima } 2\mu t_{min} = n\lambda \quad \ldots(7.6b)$$

These apply to observations made from the reflection side along the normal direction.

If the light in which the film is seen is monochromatic, (single λ), then the equations (7.6a) and (7.6b) tell us that we get alternate bright and dark fringes according to the variation of t. For a wedge, these fringes are straight lines, but in a general case, they could be quite irregular.

Fig. 7.6 Interference fringes in thin films: (a) the case of a wedge film; (b) a film of irregularly varying thickness.

If the thin film is seen in the reflected white light, where λ varies from $\sim$ 4500 A° to 7500 A°, coloured fringes appear on the film. This is because the bright fringes for the different colours fall in different positions. For instance, if at some place 2 μt equals 1 λ for red (7500 A°), it will be 1.5 λ for blue (5000 A°), so that at this place blue is best reflected, red is not reflected at all, and the intermediate colours have intermediate contributions. At another place, let 2 μt be 10,000 A° which is 2 λ for λ = 5000 A° and 1.5 λ for λ = 6667 A° and, therefore, blue is not reflected at all, while orange-red is reflected best.

It will be noted that the colours seen here are not the isolated colours, i.e., at each place there is a mixture of colours, and the composition in this mixture is different at different places. On this account, the colours as seen show very rich and impressive hues (shades of variation). However, if the film-thickness becomes large (say, $\sim$ 20 λ), then so many colours mix up that the result becomes nearly 'white' everywhere and, hence, the fringes are not seen.

EXAMPLE 7.5

A soap film held vertically and seen in the reflected white light shows red colours at the top, three intermediate red fringes and, finally, a bluish tinge at the bottom. The film then just breaks off at the top. Deduce the film-thickness at the bottom, taking $\lambda_{red} = 6500\,A°$ $\lambda_{blue} = 5000\,A°$ $\mu = 1.33$.

Solution

At the uppermost red fringe,

$$2\mu t = \left(0 + \frac{1}{2}\right)\lambda_{red},$$

where n is put as zero because the film just breaks off there. At the lowest red fringe then,

$$2\mu t' = \left(3 + \frac{1}{2}\right)\lambda_{red}$$

$$= 3.5 \times 6500\,A° = 22750\,A°$$

The bottom has a bluish tinge. Hence, there

$$2\mu t'' = \left(n + \frac{1}{2}\right)\lambda_{blue}$$

$$= \left(n + \frac{1}{2}\right) 5000\ A^\circ.$$

Since t'' is greater than t', the smallest n is 5. That gives $2 \times 1.33t'' = 5.5 \times 5000\ A^\circ$

$$t'' = 1.0 \times 10^4\ A^\circ = 1.0 \times 10^{-4}\ cm.$$

7.5 Laws of Reflection and Refraction

In the Unit on Wave-Motion, we took the experimental view of how wave fronts are reflected and refracted when they meet the boundary of two media. We gave no theoretical explanation why a given incident wave front changes direction or shape in a particular way on reflection or refraction. When the wave theory of light was first propagated, one argument against it was that waves 'spread out' and do not follow precise directions on reflection or refraction, whereas light, by and large, follows the well-known laws. Huygens was the first to give a proper theoretical explanation that waves also should, by and large, follow these laws.

Huygens' Construction

Huygens made the following two assumptions:

(i) Each point in a wave front becomes a fresh source of secondary wavelets, which spread out with the speed of light in that medium.

(ii) The new wave front at any later time is given by the forward envelope of the secondary waveletes at that time.

In Fig. 7.7, AB is the wave front at any instant t_o, spheres of the radii $c.\Delta t$ each, are drawn from several points on AB, and two envelopes CD and EF are obtained. Either CD or EF is the new wave front at the time $t_o + \Delta t$, depending on whether AB was travelling to the right or the left. It will at once be noticed that rejection of one of the envelopes is arbitrary. It will also be noticed that, in effect, all parts of the secondary wavelets are totally ignored, and only the part which lies on the forward envelope is considered useful. In a mathematical theory, both these objections are taken care of. A third point of note is that towards the ends A and B, the 'envelope' remains undefined. That is exactly where the wave model of light would differ from the ray model. In

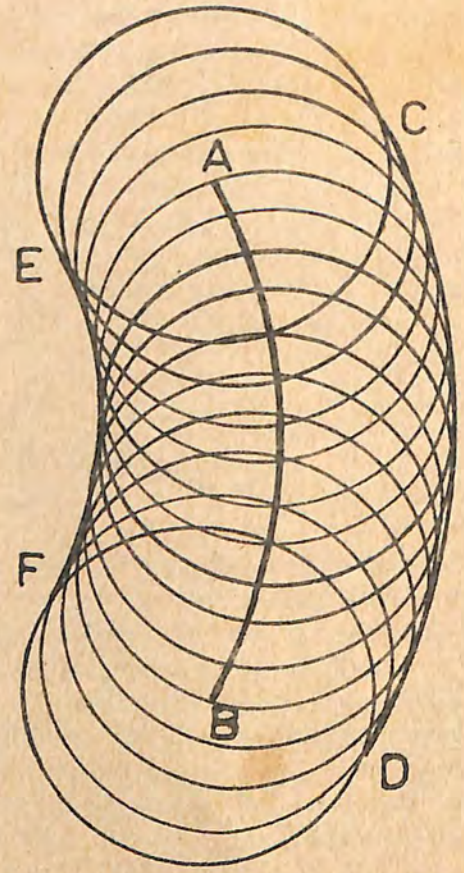


Fig. 7.7 Huygens' construction. AB is the wave front at $t = t_o$. Wavelets of the radii $c\Delta t$ are drawn from all parts of AB; the envelopes give the wave fronts at the time $t_o + \Delta t$. Is EF the new wave front or CD?

a cone of rays, the boundaries are sharp and well-defined; in a wave front, which is limited in width, the boundaries are not well-defined and sharp. Presently, we shall keep our attention to the well-defined part of the forward envelope, which is the new wavefront on Huygens' theory.

We can now use Huygens' method to deduce the laws of reflection and the laws of refraction. The second alone will be discussed as an example.

Laws of Refraction

In Fig. 7.8, let LM represent a plane surface separating two media: the velocity of light being c in the medium above it and c' in the medium below. Consider the part AB of a plane wave front travelling towards the surface LM. It will first strike at P and then successive points towards R. From each point on LM, a secondary wavelet starts growing in both media, but our attention presently will be on those in the second medium which grow at the speed c'. At the instant when R is just disturbed, the wavelet from P got time B_1R/c to grow and, hence, acquired as radius

$$PA_3 = \frac{B_1R}{c} \cdot c'. \qquad \ldots (7.7)$$

So, we draw a circle of a radius PA_3 with P as the centre to represent this wavelet. A tangent RA_3 is drawn from R to this circle. We may draw an infinity of circles of appropriate sizes from the intermediate points between P and R, and RA_3 will be a common tangent to all of them. Thus, RA_3 represents the forward envelope to the Huygens' wavelets and, hence, the refracted wavefront.

From Δs PB_1R and PA_3R, we get

$$\frac{\sin i}{\sin r} = \frac{B_1R/PR}{PA_3/PR} = \frac{B_1R}{PA_3}.$$

Using the equation (7.7), we get

$$\frac{\sin i}{\sin r} = \frac{c}{c'} \text{ a constant.} \qquad \ldots (7.8)$$

Thus, Snell's law of refraction is deduced on the wave model of light. What is of special significance is that while in ray optics the constant ratio of sines was just named refractive index μ, we now know that it is related with the wave velocities directly*.

$$\mu = \frac{c}{c'} \qquad \ldots (7.9)$$

Actually, there is another law of refraction: the refracted ray is in the plane defined by the incident ray and normal to the surface at the point of incidence. A formal proof, in brief, is as follows. In Fig. 7.8, the incident wave front AB, the dividing surface LM and the reflected wave front A_3R are all perpendi-

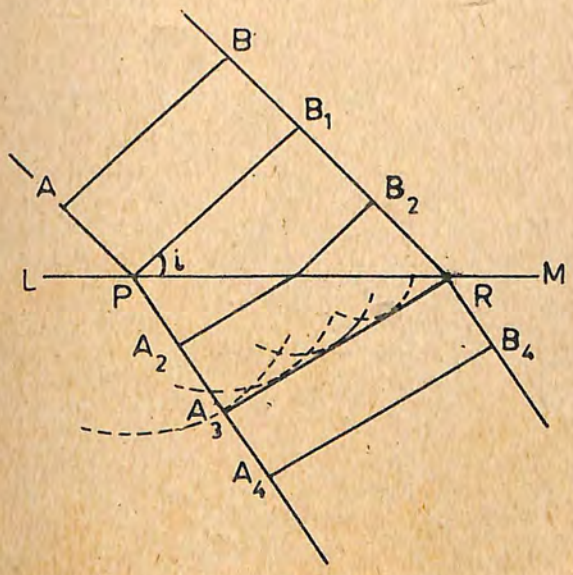


Fig. 7.8 Deduction of the laws of refraction.

* It is interesting to note that in corpuscular theory, the bending of the ray path in refraction was explained by assuming that the second medium applied an attraction force on the corpuscles. That gave c' greater than c and the theory predicted $\mu = c'/c$. Measurements of the velocity of light in various media are now available and there is quantitative evidence in favour of the result (7.9).

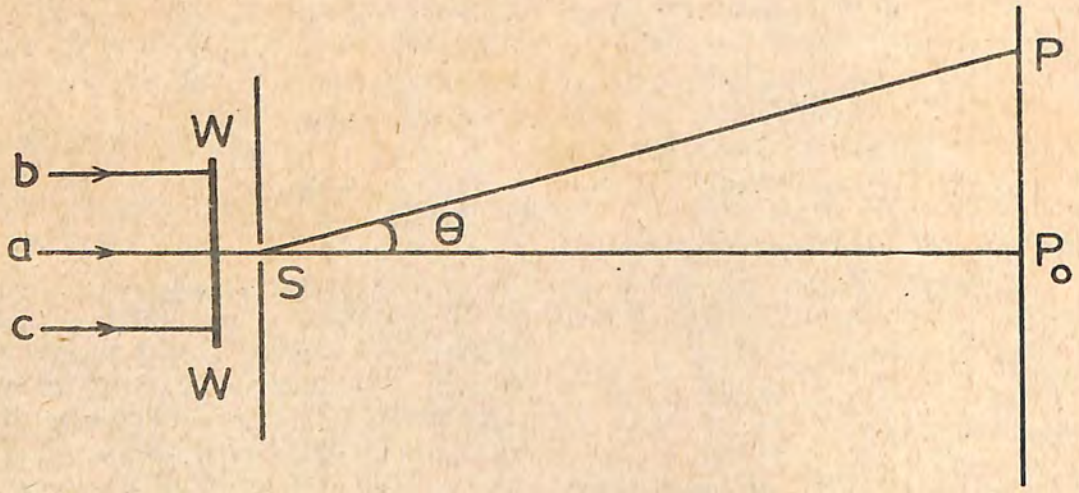


Fig. 7.9 The problem of the passage of light through a narrow slit. Will some light reach P?

cular to the plane of the paper and, hence, the incident wave normal, the normal to the surface and the reflected wave normal, all lie in the common plane, which is the plane of paper.

7.6 Diffraction of Light

We shall now discuss the propagation of light when the wave front is limited in width. Fig. 7.9 states the problem. S is a narrow slit and light from the left passes through it. In the ray model, a narrow part near *a* passes through the slit, reaching P_o, and no light would reach an off-point like P. On the wave model, the incident wave front WW disturbs each point in the slit S, and the sumtotal of the secondary wavelets from all parts of S would tell how the light goes. We proceed to examine this matter. In Fig. 7.10, we take a very much enlarged view of the slit in Fig. 7.9. The slit is now AB of the width *d*, and divided into 2 N equal parts (2 N = 12). The lines joining each of these parts with P are shown and they are parallel because the distance of P is very large compared with *d*.

The incident wave front disturbs all parts of the slit AB simultaneously. Hence, the wavelets from all 2 N parts of the slit start in the phase. But the observation point P is nearest from A and farthest from B. If AM is perpendicular from A on the lines, then the path difference BP − AP is equal to BM. If we call it *p*,

$$p = d \sin \theta. \qquad \ldots (7.10)$$

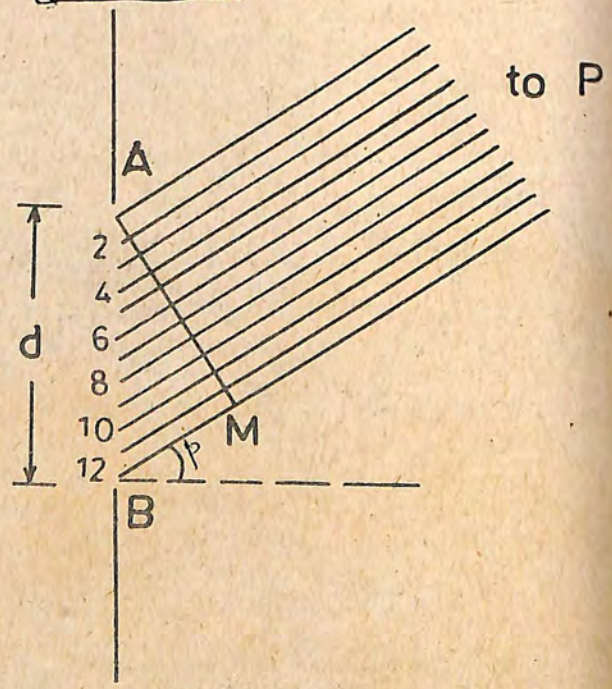


Fig. 7.10 For the theory of diffraction at a slit AB. The angle at B is θ.

Now, if the observation point P is in position P_0 of Fig. 7.9, then $\theta = 0$ and the wavelets from all parts of the slit reach there in agreement of the phase, so that the resultant disturbance has large amplitude, hence large intensity. If θ is gradually increased, the wavelets reaching P from different parts of the slit AB gradually get out of the phase. When we reach on angle $\theta = \theta_1$ such that

$$d \sin \theta_1 = \lambda. \qquad \ldots (7.11)$$

the phase difference of the wavelets from A and B when they reach P becomes 2π. If we consider the slit in two equal halves of N parts each, then the phase difference between the waves from each part 1, 2, 3 ... in one half and the corresponding part N + 1, N + 2, ... 2 N in the other half becomes π, and, hence, the total disturbance at P becomes zero. Thus, the equation (7.11) gives the angle θ on either side up to which the intensity spreads till it falls to zero.

For a typical visible light $\lambda \simeq 6 \times 10^{-5}$ cm, taking a 2.0 cm wide slit, the angular spread θ_1 will be 3×10^{-5} radian, which is too small to be noticed. But with a 2×10^{-4} cm-wide slit, the angular spread will be ~ 0.3 radian or about 20°, which cannot be missed or ignored. Actual observations show quantitative verification that (i) light spreads on both sides of the direct direction, and (ii) the spread is given by the equation (7.11). The spreading of light beyond the limits prescribed by ray optics is called *diffraction*, and light shows this behaviour just as much as the ripples discussed in Unit 6. For $d \gg \lambda$, we write the equation (7.11) as

$$\theta_1 = \lambda/d. \qquad \ldots (7.12)$$

An everyday experience of diffraction can be had if a distant lamp (100 m away) is seen through a piece of fine cloth. The lamp is seen as an enlarged disc. In fact a piece of cloth is a criss-cross of threads in mutually perpendicular directions, and therefore it shows a pattern of several weaker images of the lamp, which is exciting to see, because they rotate about the central disc as the cloth is rotated in its own plane. Presently, however, our attention is on the central disc, and it may be noted that from measured θ and λ one can estimate the spacing d of the threads in the cloth.

Limitation in Optical Instruments

Due to wave character, light shows diffraction, and due to diffraction, no image formed by even the most ideal mirror or lens can be perfect. For a point object, the image will be a disc of an angular radius λ/d. Due to this reason, the smallest angular separation which a telescope can measure is λ/d, where d is the diameter of its objective. Hence, to study the sky more accurately, we need giant telescopes. Similarly, the smallest linear separation which the best microscope can see is λ of visible light ($\sim 6 \times 10^{-5}$ cm). If we desire to examine finer details, we must use some waves which have smaller λ. An electron microscope is used to see details $\sim 10^{-7}$ cm or even less.

EXAMPLE 7.6

Estimate the largest distance from which the markings on a metre scale can be seen by (i) the naked eye, (ii) a telescope whose objective has a diameter 2.5 cm.

Solution

We take effective $\lambda = 6 \times 10^{-5}$ cm and aperture of the eye as 2 mm. The markings are of 1 mm separation, and seen from a distance D metre, the angle subtended is 1/1000 D radian. The condition required for seeing this detail is that the angular spread λ/d be less than 1/1000 D.

$$\frac{\lambda}{d} < \frac{1}{1000\,D}$$

$$\therefore D < \frac{d}{1000\,\lambda}$$

For the eye $D < \dfrac{0.2}{1000 \times 6 \times 10^{-5}}$

$\approx$ 3 metre

For the telescope $D < \dfrac{2.5}{1000 \times 6 \times 10^{-5}}$

$\approx$ 40 metre.

7.7 Lasers

In lasers, billions of excited atoms in a source are made to emit light in proper agreement with the phase and direction. This agreement is called *coherence of the source**, and it includes space coherence (i.e., direction) and time coherence (i.e., phase).

Excited atoms may give *spontaneous emission* or *induced emission* of light. In the 'spontaneous' case, the emissions by different atoms are random in directions and in phase. If the light of just the wavelength which the atoms can emit is passed through the assembly, the excited atoms may give induced emission. This light is the signal and the induced emission from each atom agrees with the signal in direction, polarization and phase. The result is that the induced wave becomes stronger and stronger as the signal passes over the successive excited atoms.

Actually, the beam has to be repeatedly reflected back and forth between two highly polished mirrors, so that the coherent radiation (i.e., the induced one) may become quite powerful in comparison with the incoherent (i.e., spontaneous) one. An important aspect is that the same light may also be absorbed by the unexcited atoms to raise them to the excited state. In any normal system, the unexcited atoms are much larger in number than the excited ones. Hence, the induced emission would not build up. If we desire the induced emission to be dominant, an essential condition to be achieved is that the excited atoms should be in excess of the unexcited ones. This is called *population inversion*.

Initially, some spontaneously emitted light acts as a signal to produce induced emission. During its successive passages, the light is both absorbed and emitted. But because of population inversion, the induced emission goes on dominating over the spontaneous one. This is called *build-up* of the laser action. In effect, the total result is that the beam becomes (i) extremely *directed one*—the angular spread limited only by the diffraction at the aperture, and (ii) *extremely coherent one*—the wavelength spread becoming extremely narrow**

Uses of Lasers

Lasers are put to many uses, depending on the properties of (i) small angular spread, and (ii) small wavelength spread. In measuring large distances by the direct pulse the reflection method—the first property is used. The distance of the moon from the earth has been measured thus. If a lens is used to focus laser light, the image size is limited by diffraction alone, so that its power may be focused on an area as small as 10^{-6} cm^2, giving sufficient intensity to melt and evaporate the material there. A laser has, therefore, been used for drilling fine holes and for cutting thick metal sheets. Its highly monochromatic character is of special significance in advanced research.

EXAMPLE 7.7

A laser has $\lambda = 7 \times 10^{-7}$ m, aperture $d = 10^{-2}$ m and sends a beam to the moon which has distance 4×10^{8} m from the earth. Deduce the angular spread of its beam and the areal spread as it reaches the moon.

Solution

For the laser, the angular spread is from

* This is distinct from the coherence between two sources as discussed in § 7.3. There, no phase relation was required between the billions of atomic emissions within each source.

** A monochromatic source of wavelength λ really has some wavelength spread $\Delta\lambda$. It is about 10^{-2} A° for the best sources. In a laser, it becomes as small as 10^{-10} A°. A laser beam is, in this sense, called extremely monochromatic.

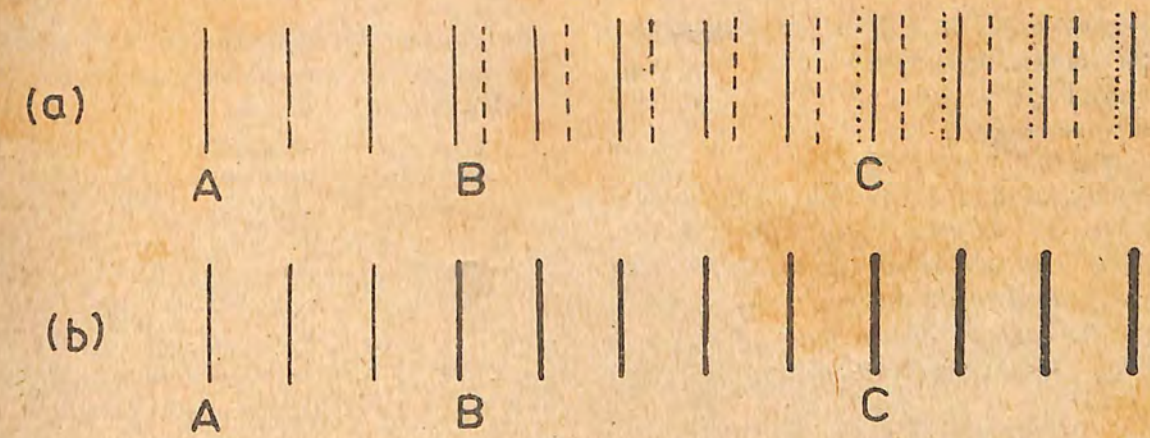


Fig. 7.11 (a) Spontaneous emissions, even if they agree in direction they do not agree in phase. Wavefronts A, those added at B and those added at C do not agree in phase (b) Induced emissions always agree in direction, and also agree in phase. Wavefronts added at B and at C just fit into those from A and hence the amplitude becomes very large when billions of such additions occur.

diffraction alone.

$$\theta_d = \frac{\lambda}{d} = \frac{7 \times 10^{-7}}{10^{-2}} = 7 \times 10^{-5} \text{ rad.}$$

The linear spread is $D.\theta_d$ and the areal spread $(D.\theta_d)^2$, where D is the distance of the moon

$$\text{Areal spread} = (4 \times 7 \times 10^3)^2$$
$$= 8 \times 10^8 \text{ m}^2$$

EXAMPLE 7.8

A laser of 10 mW power has aperture $d = 3$ mm and emits light of $\lambda = 7 \times 10^{-7}$ m. If it is focused by a lens of f = 5 cm, estimate the area and the intensity of the image.

Solution

$$\theta_d = \frac{\lambda}{d} = \frac{7 \times 10^{-7}}{3 \times 10^{-3}}$$
$$= 2.3 \times 10^{-4} \text{ rad.}$$

Areal spread is $f^2 \theta_d^2$. If we call it A,

$$A = (5 \times 2.3 \times 10^{-4})^2$$
$$= 1.3 \times 10^{-6} cm^2.$$

Intensity is power/area. If we call it I,

$$I = \frac{10 \times 10^{-3}}{1.3 \times 10^{-6} \text{ cm}^2}$$
$$= 8 \times 10^3 \text{ W/cm}^2.$$

This concentration of power would melt any metal and drill a hole through it. Notice also that the hole will be as narrow as $\sim 10^{-3}$ cm in diameter. Laser drills are now regularly used.

7.8 Spectrometer

When white light is passed through a prism, it spreads out into different colours. We say a spectrum has been formed. Since colours are now known to be related with wavelengths, we shall say that a spectrum is the wave lengthwise distribution of light from a given source. The instrument employed for the study of a spectrum is called *spectrometer*. We shall now discuss the main features of a spectrometer.

Prism Materials

In a prism, distribution occurs because the refractive index μ varies with the wavelength λ. This property is called *dispersion*.

Since we know μ is the ratio of velocity of light in vacuum (c) to that in the prism material (c'), we also say that dispersion is the wave lengthwise variation of velocity c'.

Table 7.1 gives the refractive indices of crown-glass and dense flint-glass for λ = 6563 (red) and 4047 (violet). It also gives $\Delta\mu/\Delta\lambda$ and $\Delta c'/\Delta\lambda$ values. In spectrometer work, the name dispersion refers to $\Delta\mu/\Delta\lambda$; while in higher theory work, it refers to $\Delta c'/\Delta\lambda$. Another quantity *dispersive power* is defined by

$$\omega = \frac{\mu_F - \mu_c}{\mu_D - 1} \qquad \ldots (7.13)$$

where F, C, D refer to the wavelengths 4861, 6563 and 5893 A°, respectively. It is obvious that for a spectrometer prism, one must choose dense flint-glass, not crown-glass, while for spectacle lenses, the choice has to be the opposite. For ultra-violet radiation, one uses prisms of quartz and fluoride (CaF_2), while for infra-red radiation, one uses prisms of rocksalt (NaCl), KCl, KBr, etc. The discussion below will, however, be of general character.

Minimum Deviation Condition

Fig. 7.12 shows a light beam PQ striking a prism KLM at the angle of incidence i. The beam disperses, and the extreme colours red (R) and violet (V) are shown to come out at deviations δ_R and δ_V, respectively. The angle $\delta_V - \delta_R$ between the violet and red is called *angular dispersion* by the given prism.

For a given colour, the deviation δ varies with the angle of incidence i. As i increases, the δ value at first decreases, reaches a minimum δ_m, and then starts increasing. The minimum deviation angle δ_m is related to μ and the prism angle A (Fig. 7.12) through

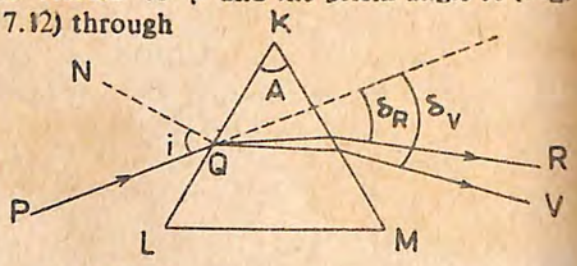


Fig. 7.12 Deviation and dispersion by a prism.

$$\mu = \frac{\sin\frac{1}{2}(A + \delta_m)}{\sin\frac{1}{2}A}. \qquad \ldots (7.14)$$

EXAMPLE 7.9

For a 60° prism, deduce δ_m for (i) μ = 1.694, (ii) μ = 1.743.

Solution

Since $\sin\frac{1}{2}(60°) = 0.5$, we get

$$\sin\tfrac{1}{2}(A + \delta_m) = \tfrac{1}{2}\mu = 0.847 \text{ and } 0.872.$$

From the tables we now get,

$$\tfrac{1}{2}(A + \delta_m) \doteq 57° 54' \text{ and } 60° 42'$$

$$\therefore \quad \delta_m = 55° 48' \text{ and } 61° 24'.$$

Spectrometer

A spectrometer has three main parts: (i) a collimator, (ii) a prism and a graduated table and (iii) a telescope. The collimator is

TABLE 7·1

The Constants for Some Optical Glasses

	μ_c λ = 6563A°	μ_F λ = 4047A°	$\Delta\mu$	$\frac{\Delta\mu}{\Delta\lambda}(A^{\circ -1})$	$\frac{\Delta c'}{\Delta\lambda}\left(\frac{ms^{-1}}{A^\circ}\right)$	ω
Crown-glass	1.5164	1.5334	.0170	-6.7×10^{-6}	$+9 \times 10^4$	.017
Dense flint-glass	1.6936	1.7427	.0491	-19×10^{-6}	$+20 \times 10^4$	.033

a tube with a convex lens at one end and a vertical slit at the other end. The slit can be slided in the tube so that it may be adjusted to lie in the focal plane of the lens.

The prism table is a circular horizontal base for the prism, rotating on a vertical axis. The horizontal axis of the collimator passes through the vertical axis of the prism table.

The telescope has its axis horizontal and passing through the axis of the prism table. It is so mounted that it can be rotated about the prism-table axis and its angular position measured on a graduated circle Fig. 7.13 gives a schematic top view of a spectrometer. S is the slit, perpendicular to the plane of paper. C is the collimator lens, its focal length equals the length SC. P is the prism. T is the telescope objective, the dotted line being its focal plane where the spectrum is formed. V and R represent the extreme colours in the visible spectrum. The eye-piece of the telescope is not shown; it is just after the plane VR. There is a vertical wire X in this focal plane, and by rotating the telescope one can bring any part of the spectrum on this wire (called a cross-wire) so as to measure δ_m for that part of the spectrum. But before this measurement, the prism P has to be rotated either way to get minimum δ_m for that part of the spectrum.

A spectrometer is usually *calibrated*, using a source for which λ for several spectral lines is known. Thus, for λ_1, λ_2, λ_3 ... we measure δ_1, δ_2, δ_3 ... and plot a graph between λ and δ. That is the calibration graph. Now, for the source under study if δ' is the minimum deviation for any part of the spectrum, the corresponding wavelength λ' can be read from the calibration curves.

Spectrographs

When the spectrum is to be photographed, the telescope (Fig. 7.13) is replaced by a camera, which really means that the eye-piece (not shown) is removed and a photographic plate is put in the focal plane RV. The instrument is then called *spectrograph*. In this case, the prism and the camera are kept fixed, and no circular scale is needed to measure angles. On the photographic plate, the observed spectrum is compared with a known spectrum, and λ values are deduced from the measurements on the plate.

The collimating and focusing work could be done by concave mirrors aluminised on the front surface. That is useful in ultra-violet and infra-red regions. Thus, spectrographs can be of many designs, but the basic principle is shown in Fig. 7.13—collimation, dispersion and focusing.

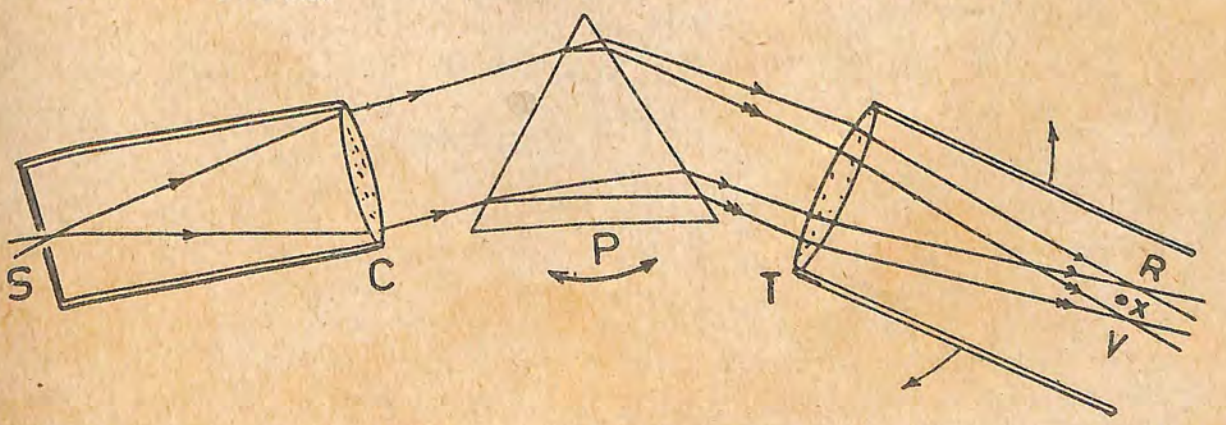


Fig. 7.13 The optical action of a spectrometer.

7.9 Various Kinds of Spectra

The spectrum of a source of light is called an *emission spectrum*. It is the wavelengthwise distribution of light emitted by that source. On the other hand, one can allow light of different wavelengths to pass through a substance and examine what fraction of light is absorbed at each wavelength. The distribution of the wave-lengthwise absorption of light by a substance is called the *absorption spectrum* of that substance. Every source has a characteristic emission spectrum and every substance has a characteristic absorption spectrum. Also if a source contains a particular substance X as an emitter, then its emission spectrum is directly related (but not identical) with the absorption spectrum of X.

Fig. 7.14 shows some typical spectra as the negatives of a photographic plate, assumed uniformly sensitive in the full range. We shall discuss them below.

Continuous Spectrum

It has radiations at all the wavelengths without break, the intensity varying continuously. Fig. 7.14a shows the case of (say) a red hot iron. The intensity is small, and becomes negligible as we reach the green region. Fig. 7.14b is the spectrum of a filament lamp temperature ~ 2000° K. The intensity increases considerably, particularly towards the blue end. Fig. 7.14 c is the solar spectrum (temperature ~ 6000° K). The maximum intensity is at yellow, and the short wavelength end (V) gains considerably relative to (b). It will be clear from the comparison of (b) and (c) why deep blue and indigo clothes appear black in the light of filament lamps. In the solar spectrum, there are, in fact, a number of dark lines, like those shown at (f). These are called *Fraunhofer lines*, which only good spectrographs are able to show.

A continuous spectrum is emitted by matter in bulk. It does not depend on the substance at all, but it depends on temperature alone. For this reason, it is also sometimes called *thermal radiation**. A hot charcoal, a hot iron, a hot filament—all emit a continuous spectrum. The sun's inner part emits a continuous spectrum corresponding to 6000° K, but the outer gaseous atmosphere absorbs some sharp lines.

Band Spectrum

In this spectrum, we observe some bright bands as in Fig. 7.14d. Each band has one edge sharp and the other end gradually fading. In fact, under a good spectrograph each band shows several separate lines, which crowd together at the sharp edge but are well spaced at the other end. (In contrast, a continuous spectrum would not show separate lines even under the best spectroscope.) The edge may be towards the violet end or towards the red end.

A band spectrum is emitted by a molecular gas, e.g., a discharge tube containing CO_2 or NH_3 or candle-light. Each molecule has its own characteristic system of bands.

Line Spectrum

In this spectrum, we observe some 'lines' as in Fig. 7.14e. These lines may be weak or intense. A line spectrum is emitted by free atoms, i.e., atoms in the gaseous state. For example, neon, sodium, mercury, etc., in discharge tubes give line spectra. Each element has its own characteristic spectrum. Some salts break up into atoms when put in a Bunsen flame, and the metal atoms give a line spectrum. Chemists use this phenomenon in the flame test to identify the metal element.

It may be noted that the mercury tube

* The term white light is also used for the light having a continuous spectrum.

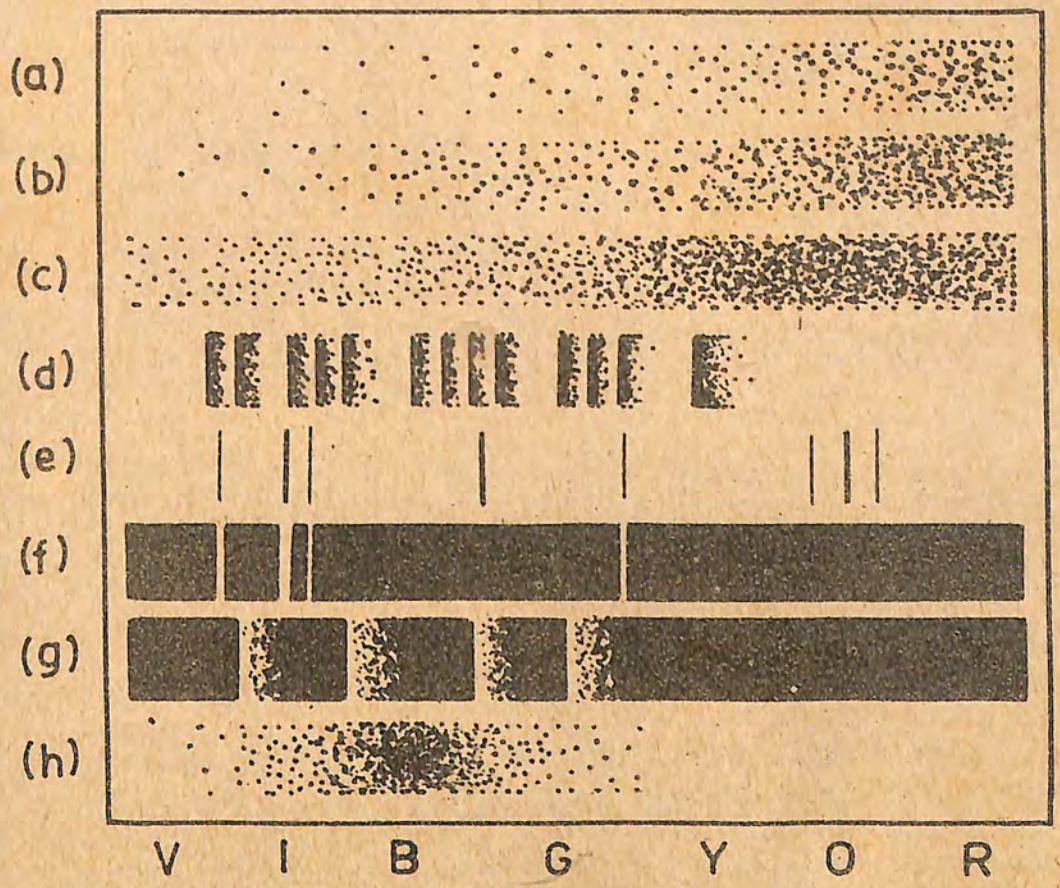


Fig. 7.14 Some typical spectra: (a) Filament emission at low temperature (~ 1000° K); (b) filament emission of a domestic lamp (temperature ~ 2000° K); (c) a sunlight spectrum; (d) a band emission spectrum of a molecular gas discharge; (e) a line emission spectrum of an element in a gas state (f) a line absorption spectrum of the same element as in (e) above; (g) a band absorption spectrum (h) a spectrum when white light is passed through a blue filter

Lights used in homes have a fluorescent coating, which gives a 'continuous' background, in addition to the line spectrum. The fluorescent material is chosen to make the wavelengthwise distribution as close to natural light (sunlight) as possible.

Line Absorption Spectrum

If white light is passed through an element in the gaseous form, the continuous spectrum has some discrete sharp lines absent, as in Fig. 7.14f. If the emission spectrum of an element X is compared with the absorption spectrum of the same element (as in Figs. 7.14e and 7.14f), all the lines in the latter are always found in the former. (The emission spectrum has several extra lines too.) Thus, if white light from a high temperature source is passed through a Bunsen flame in which asbestos soaked in NaCl is kept, then it shows a pair of dark lines in the yellow region of the otherwise continuous spectrum. They are due to Na vapour. If the source is shut off, the flame shows a pair of bright yellow lines in the position against an otherwise dark background.

As already stated, the solar spectrum shows Fraunhofer dark lines. If we search

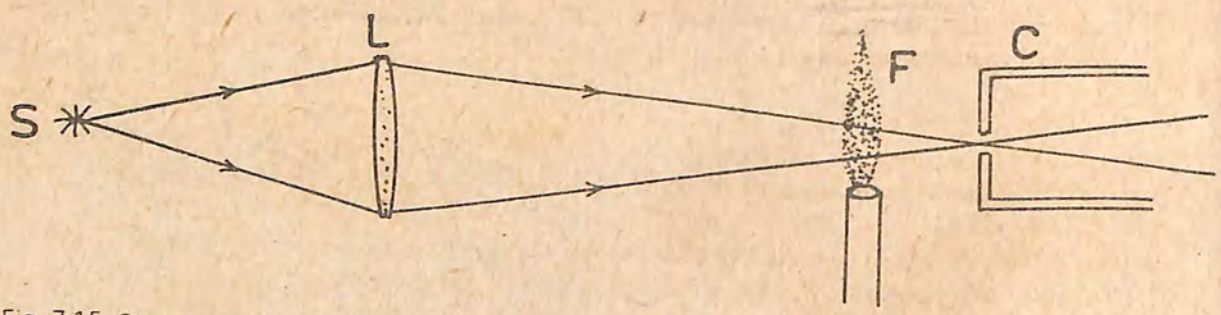


Fig. 7.15 Observing the absorption/emission spectrum of atomic sodium: S is the white light source, C is the collimator of a spectrometer, F is a Bunsen flame fed with NaCl solution. If source temperature Ts is greater than the flame temperature T_F, we get line absorption due to Na; if $T_S < T_F$, we get line emission due to Na superposed over a weak continuous spectrum.

for elements on the earth which show emission lines at the same places, we can find what elements exist in the outer atmosphere of the sun. The elements have been traced; in fact, helium was first discovered in the Fraunhofer spectrum of the sun and only later discovered on the earth.

Band Absorption Spectrum

If white light is passed through a molecular gas like I_2 vapour, the continuous spectrum has some discrete bands absent. Fig. 7.14g shows the spectrum: each absorption band has one end sharp, the other gradually fading out. The absorption bands of a given molecule do not coincide with the emission bands. Generally, one band coincides, and from there the other absorption bands lie towards the violet side, whereas the other emission bands lie towards the red side.

In a vast variety of cases, separate bands as in Fig. 7.14g may not appear in absorption, but a whole broad region of the spectrum may be absorbed. Fig. 7.14h shows a case where the whole violet region and the whole region green to red are absorbed and only a part around the blue region is transmitted. 'Filters' for light are made like this, choosing appropriate absorbing dyes.

EXERCISES

7.1 If a particle is thrown horizontally at a speed 3×10^8 m/s, deduce its vertical fall in travelling 1 km distance, given $g = 10$ m/s^2. Does the result depend on the mass of the particle? Comment on your result, considering that Newton thought light as made up of 'corpuscles' shot at a very large (infinite!) speed by the source.

(5.5×10^{-11} m)

7.2 (a) Mercury green light has a wavelength 5.5×10^{-5} cm. Deduce (i) the frequency (in Hz), (ii) the period (in sec). Convert them to megahertz and microsecond, respectively.

(5.5×10^{14} Hz ; 1.8×10^{-15} s; 5.5×10^8 M Hz ; 1.8×10^{-9} μs)

(b) In radio waves, we have a '30-metre band', which refers to a wavelength. Deduce the frequency in megahertz.

(10 M Hz)

(c) In microwaves, we may have a source of frequency 24000 Mc/s. Deduce the wavelength of the corresponding radiation.

(1.25 cm)

7.3 In Young's experiment, let lights of $\lambda = 5.4 \times 10^{-7}$ m and 6.85×10^{-8} m be used in turn, keeping the same geometry. Give a drawing like Fig. 7.2b for fringes in the two cases side by side.

7.4 Explain the meanings of (i) coherence between two sources, (ii) coherence of light from a given source. Why can two independent light sources be not coherent?

7.5 Show that in the equation

$$a_1 \cos \omega t + a_2 \cos (\omega t + \phi) = a \cos (\omega t + \phi')$$

the value of a is given by

$$a^2 = a_1^2 + a_2^2 + 2 a_1 a_2 \cos \phi.$$

Discuss the use of this for deducing the intensity when two sources of amplitudes a_1 and a_2 having the phase difference ϕ interfere.

7.6 (a) For observing interference in thin films, we need a broad source. Explain, why?

(b) Using light of $\lambda = 5.9 \times 10^{-7}$ m, it is found that in a thin film of air 7.4 fringes occur between two points. Deduce the difference of the film thickness between those points.

(2.2 microns)

7.7 Two coherent sources of intensity ratio 100 : 1 interfere. Deduce the ratio of intensity between the maxima and minima in the interference pattern.

(3 : 2)

7.8 White light may be considered to have λ from 4000 A° to 7500 A°. If an oil film has thickness 10^{-4} cm, deduce the wavelengths in the visible region for which the reflection along the normal direction will be (i) weak, (ii) strong. Take μ of the oil as 1.40.

[(i) 4000, 4667, 5600, 7000 A°; (ii) 4308, 5091, 6222 A°]

7.9 (a) The wavelength of mercury green light is 5461 A° in a vacuum. Deduce the frequency (given $c = 3.0 \times 10^8$ ms^{-1}) and the wavelength in glass ($\mu = 1.58$) and in water ($\mu = 1.34$).

(5.5×10^{14} Hz; 3460 A°; 4075 A°)

(b) In a certain region A to B in a thin film, we get 10 fringes with light of $\lambda = 4358$ A°. How many fringes would be observed in the same region with $\lambda = 5893$ A°?

(7)

7.10 (a) A slit 5.0 cm wide is irradiated with microwaves of $\lambda = 2.0$ cm. Deduce the angular spread of the central maximum, assuming incidence along the normal to the plane of the slit.

($\pm$ 23° 36′)

(b) Recalculate the angular spread on the two sides of the initial direction if the incidence is at angle 15° from the normal.

(41° 12′; −8° 6′)

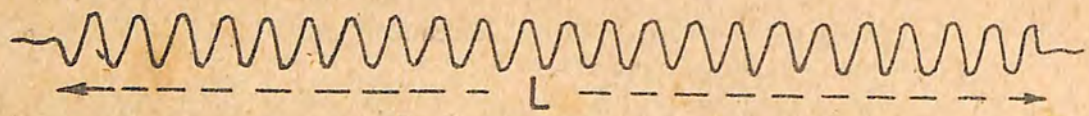


Fig. 7.16

7.11 When a distant mercury lamp is seen through a piece of fine cloth, one sees several lamps arranged in a regular rectangular pattern, intensities falling as one goes away from the central spot. Explain this. (For simplicity, assume the 'apertures' in the cloth as squares, and of the size (say) $200\lambda \times 200\lambda$.)

7.12 (a) The waves emitted by any atom or molecule must have some finite total length L (see Fig. 7.16). Can such a wave be truly represented by an expression like $\xi = a \cos(\omega t + \phi_0)$? For sodium yellow light, this length L (called the coherence length) is 2.4 cm. Deduce (i) the number of oscillations in this length, (ii) the coherence time. ($\lambda = 5.9 \times 10^{-7}$ cm, $c = 3.0 \times 10^8$ ms^{-1}).

$(4 \times 10^6; 8 \times 10^{-11}$ s)

7.13 (a) Explain the advantage of a laser beam as a 'flashlight' for measuring large distances.

(b) A laser beam is (i) highly directional, (ii) highly coherent. Explain the meaning of these two statements.

(c) Explain why a laser beam of 0.2 watt power can be focused to drill holes through a metal sheet, whereas a torch beam of even 100 watt power would not do that.

7.14 (a) From the equation (7.14), deduce δ_m for $\mu = 1.333$ (water) and $\mu = 1.740$ (methylene iodide), taking $A = 60°$

$(23° 36'; 61° 0')$

(b) For dense flint-glass $\mu = 1.642$, 1.647, 1.661 and 1.672 for $\lambda = 6.56$, 5.89, 4.86 and 4.38×10^{-5} cm, respectively. Deduce δ_m for each case (taking $A = 60°$), and plot a graph between λ and δ_m.

$(\delta = 50° 24'; 50° 52, 52° 18', 53° 26')$

7.15 (a) In Fig. 7.14 three continuous spectra are shown. Give a schematic representation of them by a graph plotting the intensity against the wavelength. Comment what differences occur as temperature of the source increases?

(b) Give a schematic representation of (i) a line absorption spectrum, and (ii) a band emission spectrum, using the I, λ curve method.

7.16 State the kind of spectrum we get from the following sources, giving a graphical I, λ representation for each case. (i) A hot iron rod, (ii) $FeCl_3$ on a platinum tip in a Bunsen flame, (iii) light from a filament lamp (T = 2200 K) passing through a Bunsen flame (T = 1400 K) in which $FeCl_3$ is held on a platinum tip, (iv) light from a wax candle, (v) a filament lamp light passed through a tube containing iodine (I_2) vapour.

7.17 What are Fraunhofer lines? How are they interpreted?

7.18 A spectrometer measures angles correct to 6' of an arc. If an experiment gives $A = 60° 0'$, $\delta_m = 48° 36'$, estimate the percentage accuracy of the value of μ. (Hint: Deduce $d\mu/d\delta_m$ expression to calculate the range of the error $\Delta\mu$.)

(0.12 per cent)

UNIT 8

Kinetic Theory of Gases

8.1 Introduction

Matter is made up of minute particles called *molecules*. On the basis of this hypothesis, many physical phenomena like the states of matter, elasticity, evaporation, etc., have been explained satisfactorily. For example, the three states of matter can be explained as follows. The molecules of any substance are held together by intermolecular forces. These forces resist the breaking up of molecular aggregates due to internal thermal agitation. If the forces due to thermal agitation are weaker than the intermolecular forces, then the substance will remain a solid and retain both its shape and volume even when its temperature is considerably raised. If the force due to thermal agitation are comparable to the intermolecular forces, then the molecules of the substance can slide past one another even at room temperature. Such substances are called *liquids*, and though they can retain their volume they take the shape of the container in which they are poured. If, on the other hand, the forces due to thermal agitation are stronger than the intermolecular forces, then the molecules of the substance fly apart and such substances are called *gases*. A gas can neither retain its volume nor its shape.

The molecules of all substances are constantly in motion. Direct experimental proof of the existence of molecular motion came from such phenomena as diffusion and Brownian motion. For example, carbon-dioxide is found to diffuse into bromine when a jar containing carbon-dioxide is placed over a jar containing bromine.

Botanist Robert Brown observed the irregular motion of plant spores floating in water. Similar motion can be easily observed, when small colloidal particles are suspended in any liquid and examined under a powerful microscope. Such irregular motion is called *Brownian motion*. A common example of this motion is the floating of smoke particles in air. Such small particles like plant spores or smoke particles act as intermediaries between macroscopic surroundings and microscopic molecules. The motion of each particle is due to the inequality of the impact of the surrounding molecules on it. These experiments on diffusion and Brownian motion support the view that the molecules of a substance are in constant motion.

The phenomena of evaporation and vaporization are intimately connected with the motion of the molecules. When a liquid is heated, the molecules move more rapidly and, as a result, some of the molecules escape out of the liquid. Such phenomena have led scientists to identify heat with molecular motion.

Based on the above two hypotheses, namely, matter is made up of molecules and heat can be identified with molecular motion, a theory has been developed to satisfactorily explain certain physical concepts, like temperature, pressure, energy, etc. Such a theory is called kinetic theory of matter. Here, only the kinetic theory of gases is considered because of its simplicity. Of course, the model thus developed should certainly explain the familiar laws like (i) gas law, (ii) Avogadro's law, and (iii) Graham's laws of diffusion, which describe the behaviour of gases.

Gas Law

What is expected to happen when a gas is compressed? Its temperature may rise, volume may decrease without a change in temperature, or both may happen. For one gram-mole of a gas, the behaviour is expressed by the gas law:

$$PV = RT \quad \ldots (8.1)$$

where V is the volume, T is the temperature, p is the pressure exerted by the gas on the walls of the vessel and R is the gas constant.

Avogadro's Law

According to this law, equal volumes of all gases, measured under the same conditions of pressure and temperature, contain the same number of molecules. Let us examine its significance. The mass of 22.4 litres of a gas at 0° C and one atmosphere pressure is equal to its molecular weight expressed in grams. For example, if the gas is hydrogen, it would weigh 2 grams; if it is oxygen, it would weigh 32 grams, etc. So, it was concluded that at 0° C and one atmosphere pressure, the gram-molecular weight of any gas must have an equal number of molecules. This number is estimated to be 6.06×10^{23}. This is known as Avogadro's number and is denoted by N. Avogadro's law also does not differentiate one gas from another.

Graham's Law of Diffusion

If two vessels containing two different gases at the same temperature and pressure are connected, then it is observed that each gas diffuses into the other. The process of diffusion appears to be active till both the vessels acquire uniform distribution of molecules. After this, there is dynamic equilibrium, i.e., the number of molecules of the two gases going out from one vessel is equal to the number of molecules entering that vessel. It is observed that different gases diffuse at different rates. The denser the gas, the slower is the rate of diffusion. This is known as *Graham's law of diffusion*, and it differentiates one gas from the other.

Gas Model

A model helps in explaining and understanding a large body of facts, both qualitatively and quantitatively. The model should be quantitative. What should be the model of a gas which could explain all the above laws?

The physical quantities which describe the behaviour of a gas are pressure, volume, and temperature. Pressure is force per unit area, force is proportional to the rate of change of momentum and momentum is the mass multiplied by velocity. The mass of a body is the sum of the mass of its constituent particles, molecules or atoms. Hence, on simplification it should be possible to express pressure in terms of the number of particles, their mass, their velocities and the area of the surface experiencing pressure. Volume can be expressed as the product of three lengths, and temperature is expressed directly. So, the model of a gas explaining its behaviour should include quantities, like the mass of particles, the number of particles and their velocities, volume and temperature.

8.2 Assumptions for the Development of the Kinetic Theory of Gases

The kinetic theory of gases is based on the following simplifying assumptions:

(i) A gas consists of molecules and these molecules are perfectly elastic spheres. This implies that molecules do not deform during a collision and that the time interval during which the collision takes place is negligible compared with the time between two collisions. So, energy is not dissipated and the kinetic

energy will remain conserved during the collision.

(ii) Molecules are in random motion, i.e., they move in all directions with all possible velocities, and do not have any preferred direction or position.

(iii) No appreciable force acts on molecules, except during collisions. Forces like mutual attraction or repulsion between the molecules or between the molecules and the walls of an enclosure are negligible. This implies that the energy is totally kinetic.

(iv) The size of a molecule is negligible compared with the average distance between two molecules.

These simplifying assumptions help to avoid mathematical complications. They also assure that the model developed will be applicable to all the gases, whether it is H_2 or CO_2, etc. But, strictly speaking, none of these assumption is true. They only set a limit beyond which the model developed is not applicable. These limitations become apparent when the theory is generalized and the region of application is broadened. For example, the model thus developed does not satisfactorily explain the behaviour of liquified gases. However, the model thus developed, though approximate, satisfactorily explains a large number of experimental facts about gases.

8.3 Expression for the Pressure Exerted by a Gas

Consider one gram molecule of a gas in a cubical vessel whose walls are perfectly elastic and whose sides are of the length L. The molecules are in random motion. They collide with one another and also with the walls of the vessel. Since a large number of molecules impinge on the walls, the walls experience a force. The force experienced per unit area is the pressure exerted by the gas on the walls.

Consider a molecule moving with a velocity C. Let its mass be m. Its velocity can be resolved into three components u, v, w in the direction of the edges of the cube (Fig. 8.1) along the x-, y- and z-axis such that

$$C^2 = u^2 + v^2 + w^2. \qquad \ldots (8.2)$$

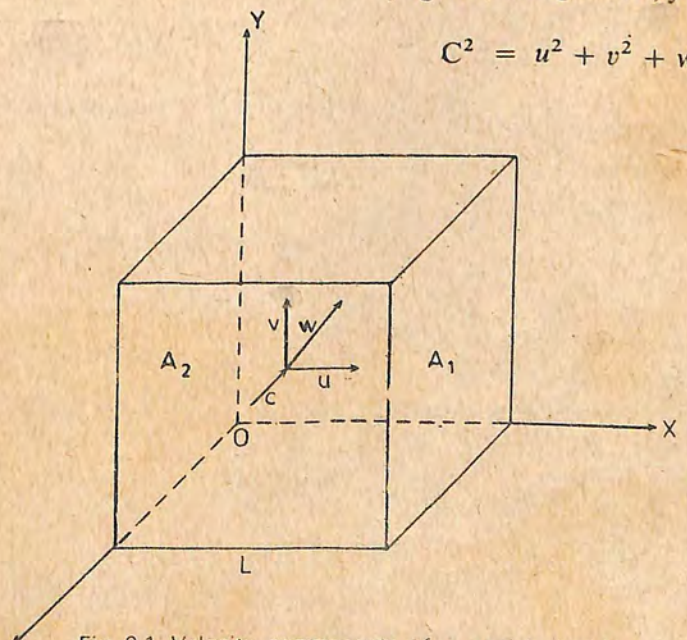


Fig. 8.1 Velocity components of a particle inside a vessel.

Let A_1 and A_2 be the faces of the cube which are normal to the x-axis. If the molecule collides with the side A_1, it will rebound with a velocity equal to $-u$. Its components v and w will remain unchanged.

The change in the momentum of the molecule on account of this collision will be $= m(-u) - mu$.

How much momentum is imparted to the wall A_1 in this process? It is well known that the total momentum has to be conserved. So, the momentum imparted to A_1 should be 2 mu. What will be the momentum imparted to the wall A_1 in unit time interval?

The molecule after rebounding from A_1 will return and collide with the wall A_2. The time it takes to move A_2 from A_1 will be L/u.

So, after every interval time of 2 L/u, it will collide with the wall A_1. So, in unit time, the number of collisions it will make with the wall A_1 will be u/2 L. So, the momentum imparted to the wall A_1 in unit interval of time is

$$(2\,mu)\,\frac{(u)}{2L} = \frac{mu^2}{L}. \qquad \ldots (8.3)$$

By Newton's second law, it follows that the force exerted on the wall A_1 by the molecule will be $\frac{mu^2}{L}$.

Since pressure is defined as force per unit area, the pressure exerted by the molecule on A_1 will be

$$mu^2/L^3. \qquad \ldots (8.4)$$

If the vessel contains N molecules, then the pressure exerted by them on the wall A_1 will be

$$p = \frac{m}{L^3}(u_1^2 + u_2^2 + \ldots + u_N^2) \qquad \ldots (8.5)$$

$$= \frac{m}{V} N\overline{u^2} \qquad \ldots (8.6)$$

where $V = L^3$ is the volume of the vessel and $\overline{u^2}$ is the average value of u^2 of all the N molecules.

Since the molecules move entirely at random, it follows from the equation (8.2) that

$$\overline{u^2} = \overline{v^2} = \overline{w^2} = \frac{1}{3}\overline{C^2}.$$

Substituting this in the above equation, we get

$$p = \frac{1}{3}\frac{mN\overline{C^2}}{V} \qquad \ldots (8.7)$$

Since the pressure exerted by the gas is the same in all directions, it follows that the pressure exerted by the gas in any direction is

$$p = \frac{1}{3}\frac{mN\overline{C^2}}{V}$$

$$= \frac{1}{3}\rho\,\overline{C^2} \qquad \ldots (8.8)$$

when $\rho = \frac{mN}{V}$ denotes the density of the gas (mass of the molecules in unit volume of the gas).

Using the equation 8.8, root mean square speed, C_{rms}, of any gas can be easily calculated. For example, in the case of air,

$$\rho = 1.293 \times 10^{-3}$$

$$\text{and } C_{rms} = \sqrt{\overline{C^2}}$$

$$= \sqrt{\frac{3p}{\rho}}$$

$$= \sqrt{\frac{3 \times 76 \times 13.59 \times 981}{0.001293}}$$

$$= 0.485 \text{ km/s}.$$

This shows that the speed of the air molecules is comparable to the speed of the sound in air which is about 0.33 km/s.

The kinetic theory predicts the velocity

of the air molecules to be of the order of 0.5 km/s. This is much higher than the speed at which the smell of a scent appears to spread in a room. How do we explain it? The guess is that a molecule undergoes multiple collisions back and forth, tracing out a zig-zag path and, as a result, effective displacement per unit time is much less.

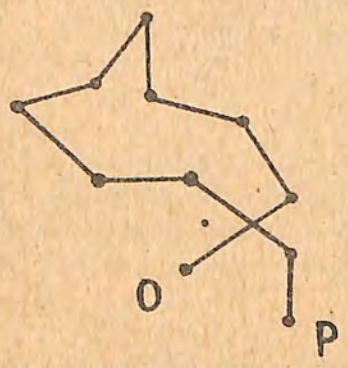


Fig. 8.2 The zig-zig path of a molecule after several collisions

For example, a molecule starting its journey from 0 may end up at P after several collisions (Fig. 8.2). It is reported, in the case of bromine diffusing into air, the bromine gas molecule in travelling an effective distance of 0.1 m may have enroute 10^{12} collisions.

8.4 Deduction of the Laws

By making certain assumptions in Art. (8.2), we have developed the kinetic theory of gases which includes quantities like the mass, velocity, number, temperature, etc., of the molecules. Let us now examine how far this theory explains the various laws which describe the behaviour of gases.

Boyles' Law

We have from the equation (8.7).

$$pV = \frac{1}{3} mN\overline{C^2}$$

Since both m and N are constants, it follows that pV will remain a constant as long as $\overline{C^2}$ is a constant. We know that if temperature remains constant, there is no flow of heat energy. So, in our system in which the energy is supposed to be totally kinetic, there will be no change in its energy, or, the mean square velocity of the gas molecules will remain a constant. This then leads us to Boyles' law, that at constant temperature the product of the pressure and volume is a constant.

Avogadro's Law

Consider two gases *a* and *b* which are at the same pressure and temperature. Then from the equation (8.7)

$$p = \frac{1}{3} \frac{m_a N_a \overline{C_a^2}}{V_a}$$

$$\text{and } p = \frac{1}{3} m_b \frac{N_b}{V_b} \overline{C_b^2}.$$

Since the two gases have the same temperature, the average kinetic energy of any molecule of the gas *a* will be equal to the average kinetic energy of any molecule of the gas *b*.

$$\frac{1}{2} m_a \overline{C_a^2} = \frac{1}{2} m_b \overline{C_b^2}$$

Substituting this in the above equations, we get

$$\frac{N_a}{V_a} = \frac{N_b}{V_b}.$$

This means all gases at the same temperature and pressure contain an equal number of molecules in equal volumes. This is Avogodros' law described in § 8.1

Graham's Law

Consider the gas *a* diffusing into the gas *b*. When the pressure exerted by th

two gases become equal to one another, it follows from the equation (8.8) that when $p_a = p_b$,

$$\frac{1}{3}\rho_a \overline{C_a^2} = \frac{1}{3}\rho_b \overline{C_b^2}$$

$$\text{or} \quad \frac{C_{a\,rms}}{C_{b\,rms}} = \sqrt{\frac{\rho_b}{\rho_a}}. \qquad (8.9)$$

This implies that gases diffuse at different rates, the denser the gas, the slower is the rate of diffusion. This is Graham's law of diffusion (8.1).

Thus, we see that the model we have developed satisfactorily explains the laws describing the behaviour of gases.

8.5 Relation between Temperature and Kinetic Energy

Consider one gram molecule of a gas. Let it have N molecules. The total kinetic energy of translation of the N molecules in the vessel can be expressed as

$$E = \frac{1}{2} mN\overline{C^2}. \qquad \ldots (8.10)$$

From the equation (8.7), we have

$$pV = \frac{1}{3} mN\overline{C^2}.$$

Using the gas law, we get

$$pV = RT = \frac{1}{3} mN\overline{C^2}$$

Substituting this in the equation (8.10), we get

$$pV = RT = \frac{2}{3}\left(\frac{1}{2} mN\overline{C^2}\right)$$

$$= \frac{2}{3} E.$$

$$\text{or} \quad E = \frac{3}{2} RT. \qquad \ldots (8.11)$$

It is sometimes useful to express this relation using Boltzman's constant $k = R/N$ which denotes the gas constant for a single molecule. From the equation (8.11), we get

$$E = \frac{3}{2} kNT. \qquad \ldots (8.12)$$

or the mean kinetic energy of translation of one molecule will be

$$E/N = 3/2\, kT. \qquad \ldots (8.13)$$

These relations clearly show that temperature is intimately related to the kinetic energy of a molecule. Whenever a gas gains energy, its temperature rises.

8.6 The Law of Equipartition of Energy

What do we mean by degrees of freedom while discussing the thermal state of a system?

In our gas model, the linear velocity C of a monoatomic molecule like helium in translatory motion was represented by the three components u, v and w along the three perpendicular directions. These components are independent of one another. The kinetic energy of a gas molecule will be $1/2\ mu^2 + 1/2\ mv^2 + 1/2\ mw^2$, and by our assumption that the molecule has only kinetic energy, this expression also gives the energy of the molecule. Since there are three independent squared terms in the expression for its energy, the molecule is said to have three degrees of freedom.

We have seen (equation 8.11) that the mean energy of a gram-molecule of a monoatomic gas is 3/2 RT. Since the molecules have three degrees of freedom and $u^2 = v^2 = w^2 = 1/3\ C^2$, it follows that the energy associated with each degree of freedom is 1/2 RT.

Since there are N molecules in a gram-molecule, it follows that the mean kinetic energy associated with each degree of freedom of any molecule is 1/2 kT.

The fact that energy is equally divided between the various degrees of freedom is known as the *law of equipartition of energy*. From this it follows that the energy of a diatomic molecule which has five degrees of freedom is 5/2 kT, and that of a triatomic molecule which has six degrees of freedom is 3 kT, etc.

8.7 Specific Heat of Gases (C_p and C_v)

Let us obtain an expression for the specific heat of a monoatomic gas such as helium. In a monoatomic gas, the molecules are identical with atoms. It is assumed that the atoms possess only kinetic energy. Therefore, the total energy E in a gram-molecule of one gas is:

$$E = \frac{3}{2}RT.$$

Since heat has been identified with molecular motion, this expression can be used to obtain the heat lost or gained due to the change in temperature. The molar specific heat of a gas at constant volume C_v is defined as the quantity of heat required to raise the temperature of one gram-molecule of a gas, maintained at constant volume, by one degree, i.e.,

$$C_V = E_{T+1} - E_T = \frac{3}{2}R.$$

Suppose, the temperature of one gram-molecule of a gas is raised from T to T + 1 at constant pressure In doing so, say, the volume of the gas changes from V to V', i.e., the gas does some external work p (V' − V). [It may be noted that the work done = (F) (d) = (p) (A) (d) = p (v)]. Therefore, for the same temperature rise, at constant pressure an extra quantity of energy equal to the external work done by the gas will have to be supplied. So,

$$C_p - C_v = \text{external work done}$$

$$= p(V' - V)\ \frac{\text{joules}}{\text{mole °k}} \qquad \ldots (8.14)$$

According to the gas law,

$pV = RT$ before heating

$pV' = R(T + 1)$ after heating

or $p(V' - V) = R$.

Hence,

$$C_p - C_v = R\ \frac{\text{joules}}{\text{mole °k}} \qquad \ldots (8.15)$$

Converting work units into heat units, we get

$$C_p - C_v = \frac{R}{J}\ \text{cal/deg} \qquad \ldots (8.16)$$

where $J = (4.18)(10^7)$ ergs/mole °k.

For a monoatomic gas, since $C_v = 3/2\ R$, C_p will be equal to 5/2 R.

From the above discussion, we see that the simple kinetic model of gases has explained the behaviour of gases within a reasonable range of temperature and pressure. However, it gives no idea about the velocity distribution, i.e., the number of molecules moving with one velocity, with another velocity, and so on. Besides this, the model will be applicable only to perfect gases, i.e., the gases obeying Boyle's law. The model, therefore, needs improvement regarding the size of the molecules, the forces between them, etc., and also from the point of view of mathematical techniques. These considerations lead to Van der Waal's equation of state, and Maxwell Boltzmann's distribution law, both of which are beyond the present scope of treatment.

EXERCISES

8.1 "The Brownian motion of a colloidal particle is due to inequality of the impact of the surrounding molecules of the medium". Briefly explain this statement.

8.2 Explain the phenomenon of evaporation.

8.3 Briefly explain why there is practically no atmosphere on the surface of the moon.

8.4 Using the ideal gas equation, determine the value of R (one gram-molecule of a gas at S.T.P. occupies 22.4 litres).

(8.3 joules/mole/K)

8.5 If three molecules have velocities 0.5, 1 and 2 km/s, respectively, calculate the relation of the root mean square speed and the average speed.

(2:1)

8.6 (a) Calculate (i) the root mean square speed and (ii) the mean kinetic energy of one gram-molecule of hydrogen at STP (given that the density of hydrogen is 0.09 kg/m^3).

(b) Given that the mass of a molecule of hydrogen is 3.34×10^{-27} kg, calculate Avogadro's number.

(c) Calculate Boltzman's constant k.

(1.38×10^{-23} joules/molecule/K)

8.7 Our atmospheric air is a mixture of gases such as hydrogen, oxygen, carbon-dioxide, etc. Using the kinetic theory of gases, prove that the total pressure exerted by the air is the sum of the partial pressures exerted by hydrogen, oxygen, carbon-dioxide, etc. (Dalton's law of partial pressures), i.e.,

$$P = P_1 + P_2 + P_3$$

8.8 Calculate C_p and C_v for Argon (given R = 9.3 joules/mole/K)

8.9 The mass of air inside a room $3 \times 4 \times 6$ cubic metres can be compared to the weight of a

A. Pin
B. Pencil
C. Table
D. Truck

(a) Tick the suitable answer.

(b) Compare your answer by calculation. (Given that the density of air is 1.3 kg/m^3).

(C)

UNIT 9

Physics of the Atom

9.1 The Nature of Matter

A question, whether matter could be subdivided without limit or not, was raised quite early in human civilization. The idea that matter composed of small indivisible particles was first conceived by an Indian philosopher and saint, Kanada, in B.C. 600; he named them *paramanu*. About a century later, a Greek philosopher, Democritus, visualized the minutest indivisible particles of matter and named them atoms.

Dalton applied the atomic theory of matter to explain the laws of chemical combination and decomposition of substances.

The laws of electrolysis by Faraday established the electrical nature of matter; matter is composed of positive and negative charges.

It was the studies on the conduction of electric charges through gases at rarefied pressures that established that atoms could have structures; these studies first resulted in the discovery of cathode rays. The structure of an atom and its interaction with electromagnetic radiation is the main content of this Unit and we shall discuss it with more stress on physical ideas rather than on mathematical details.

9.2 Cathode Rays

The conduction of electric charges through gases resulted in several discoveries. The equipment used in such studies is shown in Fig. 9.1. Two electrodes—a cathode and an anode—enclosed in a glass-tube are connected to a high voltage of 10,000 volts. A side tube is connected to a pump so as to maintain the air pressure inside the tube at the required level. When a high voltage is switched on, a small current flows through the discharge tube, this signifies a flow of charges between the electrodes. By conven-

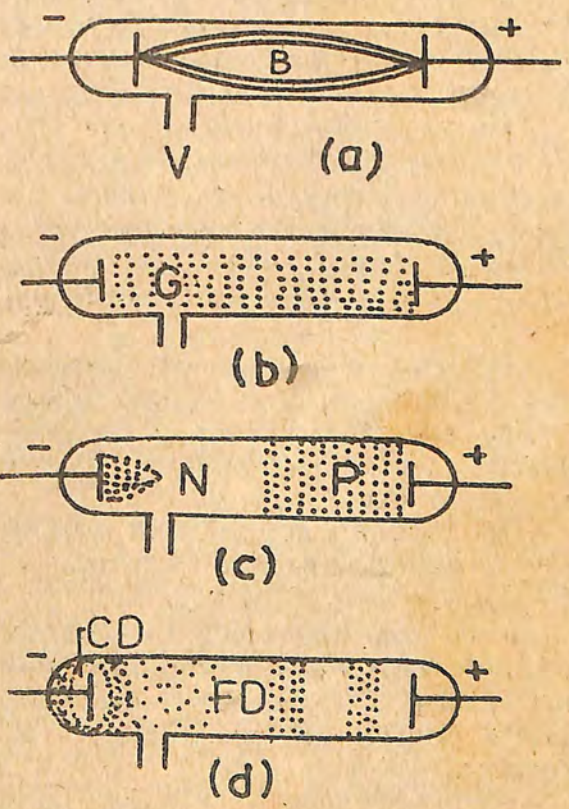


Fig. 9.1 The general appearance of the electrical discharge at various pressures: (a) At 10mm Hg: V-side tube, B− electric discharge.
(b) At 5mm Hg: G−Guisser discharge (pink).
(c) At 2mm Hg: N−Negative glow (blue); p−positive column, (pink).
(d) At pressure < 1mm Hg: CD−crookes dark space; FD−Faraday dark space.

tion, negative charges flow towards the anode (+) and positive charges towards the cathode (−). Various processes that occur in the discharge tube, at different pressures (10 mm to < 1 mm of mercury), are illustrated in Fig. 9.1. At very low pressure $\lesssim 10^{-4}$ mm of Hg, the inside of the discharge tube becomes dark and a greenish glow appears on the walls of the tube near the anode. The air inside the tube is so rarefied that charges flow from one electrode to the other. A small perforation in the anode produces a greenish spot on the glass behind the anode. This glow is due to the new rays called *cathode rays*, which impinge on the glass wall. The cathode rays originate from the cathode and flow towards the positive anode; thus, the cathode rays must carry negative charges.

The mica vanes of a wheel rotate when placed in the path of the cathode rays. This experiment shows that the cathode rays impart momentum and energy to the material objects placed in their path. Thus, mass and velocity could be attributed to the cathode rays.

The effect of an electric field perpendicular to the path of the cathode rays is shown in Fig. 9.2b. The direction of deflection shows that the cathode rays are the negatively charged particles.

The effect of a magnetic field, which is in a direction perpendicular to the plane of the paper, is shown in Fig. 9.2c.

From these experiments, it is clear that a cathode ray is a particle with the mass m and a charge e. The charge to mass ratio for the cathode rays was determined by Thomson.

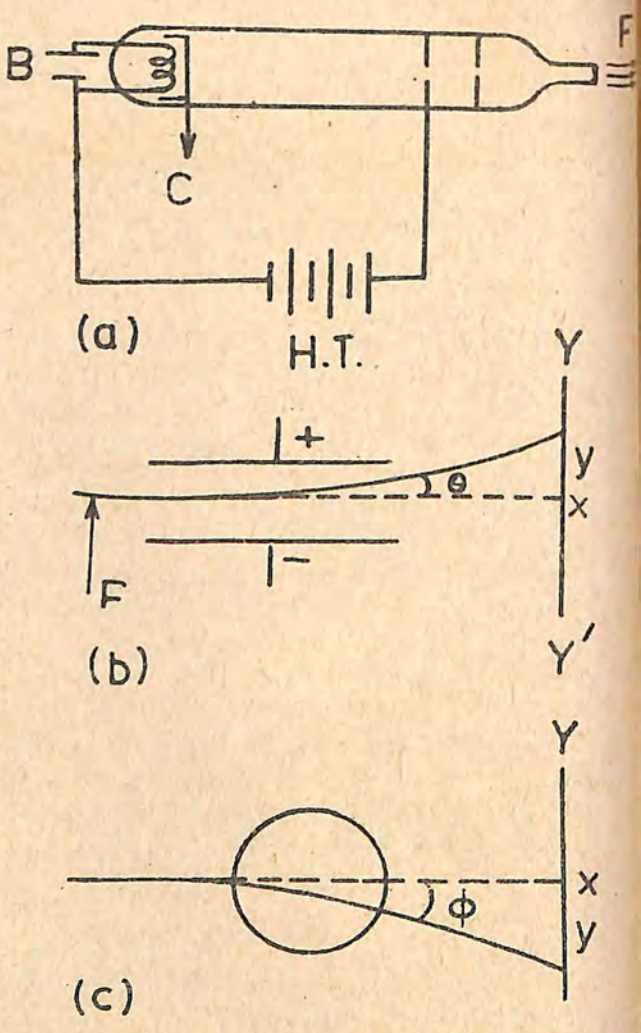


Fig. 9.2 (a) A cathode ray gun: B—battery; C—Cathode F—fine beam of electrons; H.T.—high tension.
(b) An electron beam in an electric field: YY' = Zinc sulphide screen.
(c) An electron beam in a magnetic field.

Thomson's Experiment for e/m

If V is the potential difference in volts between the electrodes of a cathode ray tube (Fig. 9.2a), then the energy gained (ev) by the cathode rays in moving from a cathode to an anode appears as kinetic energy of each cathode particle. Thus, we have

$$eV = \frac{1}{2} mu^2 \qquad \ldots (9.1)$$

where e is in coulombs, m in kg and u in m/s. From this expression, the velocity of a cathode ray is determined.

In Fig. 9.2b, the effect of an electric field E (in newtons/coul.) on the cathode rays is considered. Since E is perpendicular to the motion of the cathode rays, the negative particles experience a force (eE) along the

positive y-direction as shown in the figure. The deflected beam, after emerging from the field of the plates, moves in a straight path. The beam is incident on a screen at a distance y from its undeflected position.

The effect of a magnetic field (B webers/m^2) on negative ions is shown in Fig. 9.2c. The negative particles, which move with a velocity u, experience a force Beu along the negative y-direction*.

If we apply electric and magnetic fields at the same time over the same region on the cathode rays, the net force is along the y-axis. The beam will be undeflected if the forces by the electric and magnetic fields are counter-balanced, i.e.,

$$eE = Beu \qquad \ldots (9.2)$$

The velocity of the cathode rays from the above equation is $u = E/B$. But from the equation (9.1), we know the velocity, and combining the two equations, we obtain

$$\frac{e}{m} = E^2/2VB^2 \qquad \ldots (9.3)$$

Thus, knowing the values for E, V and B for no deflection of the beam condition in the experiment, the value of e/m for the cathode rays is determined. The value of e/m for the cathode rays was determined by Thomson, not exactly in the way described above but the principle of the method was the same. It turned out that the cathode rays from any material are identical (same e/m) and they are now known as electrons. The nature of the cathode rays is independent of the coatings on the cathodes and the gas of the electron tube. An important conclusion emerges, i.e., electrons are fundamental constituents of all materials. The current accepted value of e/m is 1.76×10^{11} coul./kg.

Millikan's Oil Drop Experiment

The experiment is designed to determine the charge on electrons. The experimental arrangement is shown in Fig. 9.3. Two parallel metal plates are connected to a high tension B through a switch S. In the on-position of the switch S, the plates are charged and an electric field E is established between the plates. In the off-position of S, the plates are at zero potential of the earth. The upper circular plate has a small hole at its centre. The space between the plates is illuminated by a source L and a converging lens. The circle represents the field between the plates as viewed through a microscope. A fine shower of oil is sprayed by means of an atom sprayer on the upper plate. Fine drops of oil enter through P and are viewed through the microscope. Normally, by friction, some charge is produced on an oil drop. With the switch in the off-position, the plates are at zero potential and the oil drop descends towards the lower plate under the action of the gravitational force. Since the drop is very small (say, of the mass M) its descent is opposed by air friction Fig. 9.3 (c)– and the drop very soon achieves a terminal constant velocity V_o. The velocity V_o is measured by recording the descent of the drop in a given time interval Fig. 9 3 (b). The terminal velocity enables one to determine the mass M of the oil drop. Before the drop reaches the bottom plate, the switch S is brought to the on-position. Since the oil drop is negatively charged, it moves upward towards the positively charged plate with a terminal velocity V_1 which again is measured. If the electric intensity E is so much that its pull upwards is balanced by the gravitational downward pull, then the drop remains stationary in the microscope's field of view. The situation is shown in Fig. 9.3 (d) where

* The force on a charged particle by an electric field **E** and a magnetic field **B** is given by Lorentz force. The expression is $\mathbf{F} = q(\mathbf{E} + \mathbf{u} \times \mathbf{B})$. where q and u are the charge and the velocity, respectively, of the particle.

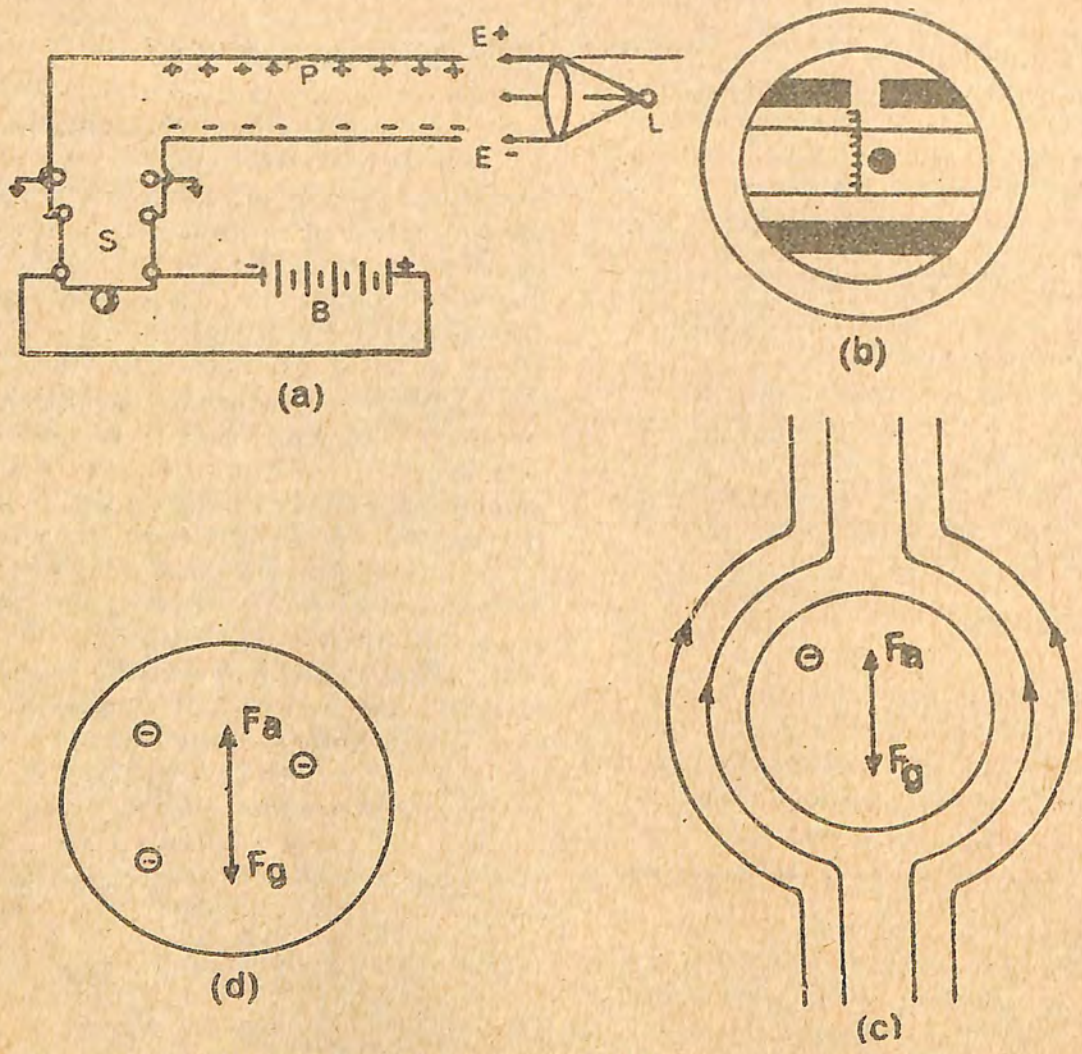


Fig. 9.3 (a) A device used in Millikan's oil drop experiment.
(b) An oil drop as seen in microscope.
(c) Gravity and air friction forces on an oil drop.
(d) Electric and gravitational forces on an oil drop.

the drop is supposed to carry three electrons. For the equilibrium of the drop, we have

$$Mg = neE \qquad \ldots (9.4)$$

where ne represents the total charge on the oil drop. The knowledge of M, E and g enables us to calculate ne. The experiment was repeated several times and the result was that the charge on the oil drop always appeared as an integral multiple of e. The current accepted value of e is 1.60×10^{-19} coulombs. The sign of the charge on an electron is negative by convention. This experiment established the quantisation of charges in nature, i.e, charges occur in integral multiples of e.

From the values of e/m and e, the mass of an electron is 9.11×10^{-31} kg.

Canal Rays

If the electrons move from a cathode to an anode in a discharge tube, it is natural that positive charges must move from the anode to the cathode. These positive charges

are called *canal* or *positive rays*. Fig. 9.4 illustrates their production. The e/m ratio of the canal rays is not unique for all elements. It has been found that most mass of the atoms is with positive rays. Wien identified rays of hydrogen with protons whose

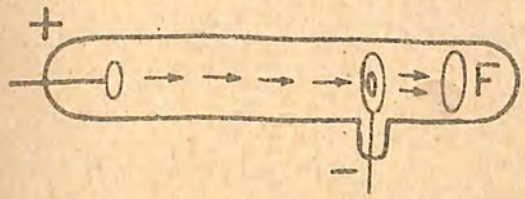


Fig. 9.4 A sketch showing the production of canal rays. → canal rays. F = fluorescent screen.

charge and mass, respectively, are 1.60×10^{-19} coul. and 1.67×10^{-27} kg.

9.3 Model of an Atom

From the above results, it emerges that an atom of any element is composed of electrons and positive charges; an atom as a whole is neutral and stable. The total positive and negative charges balance exactly. It is evident that an atom has a structure.

Rutherford's Model of an Atom

An important step forward for structural studies of atoms was presented by Rutherford. This model was based on experiments—the scattering of α-particles (alpha particles) from thin foils—performed by Geiger and Marsden.

An α-particle is a helium atom from which two electrons are removed. This is the doubly ionised state of a helium atom. Its charge is +2e and its mass is nearly four times the mass of a proton.

A schematic diagram for scattering of α-particles from a thin foil is shown in Fig. 9.5. A narrow beam of monoenergetic α-particles is incident on a screen T. In the absence of the thin foil B, the beam travels along the dotted line and impinges on the screen at D. With the foil inserted in the path of the beam, the

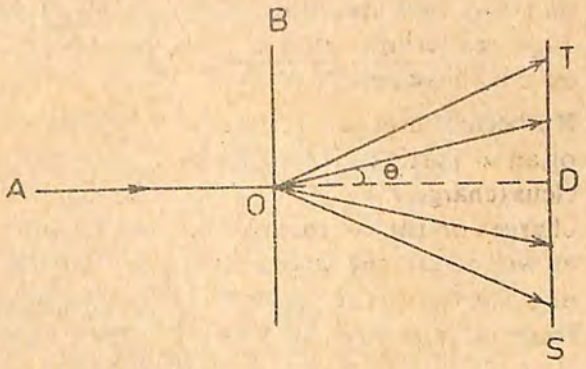


Fig. 9.5 α-scattering experiment from thin foils. AO = α-particle beam, B = thin foil, S = Screen.

individual particles get scattered and are incident at different positions on the screen. The amount of deviation of an α-particle from its original direction defines its scattering angle. This experiment has to be conducted inside an evacuated chamber, for, otherwise, the α-particles get scattered from the atoms of air also. A scanning of the positions of the α-particles on the screen determines the scattering angles (θ's) of the α-particles with respect to the line AD. The results of Geiger and Marsden are shown in Fig. 9.6. The squares are experimental points. The number $N(\theta)$ of the scattered α-particles of the angle larger than θ *versus* the angle of scattering θ is plotted in the figure. Different scales for $N(\theta)$ are chosen for clarification of the data. The following conclusions emerge from Fig. 9.6. The bulk of the α-particles is scattered by small angles. Only one particle in ~ 9000 is scattered by an angle larger than 90°. In a very few cases, the α-particles retrace their incident paths, Fig. 9.7.

The scattering of an α-particle is due to the coulomb interaction of the α-particle with positive charges and electrons present in an atom of the foil. If positive charges and electrons of an atom are distributed uniformly over its whole volume (radius ~ 10 m^{-10}), then the α-particle scattering angle θ is very small and the possibility for a large angle (> 90°) scattering is negligible (1 in 10^{12} particles). Thus, the experimental observa-

tion of Geiger and Marsden, especially large angle scatterings, could not be accounted for on the above picture of the atom.

Rutherford argued that large angle scattering of an α- particle (charge Z) by a target nucleus (charge Z') is possible when the positive charges on the α— particle (helium nucleus) as well as on, the target nucleus are treated as point charges approximately, for then the force of repulsion between the charges is given by Coulomb's law $F = \frac{ZO'e^2}{4\pi r^2}$. In otherwords, the size of the nucleus can be estimated by the study of the large angle scattering of α— particle by heavy elements. Clearly the volume of the nucleus must be much smaller than the size of the atom for the above ideas to be valid. These were

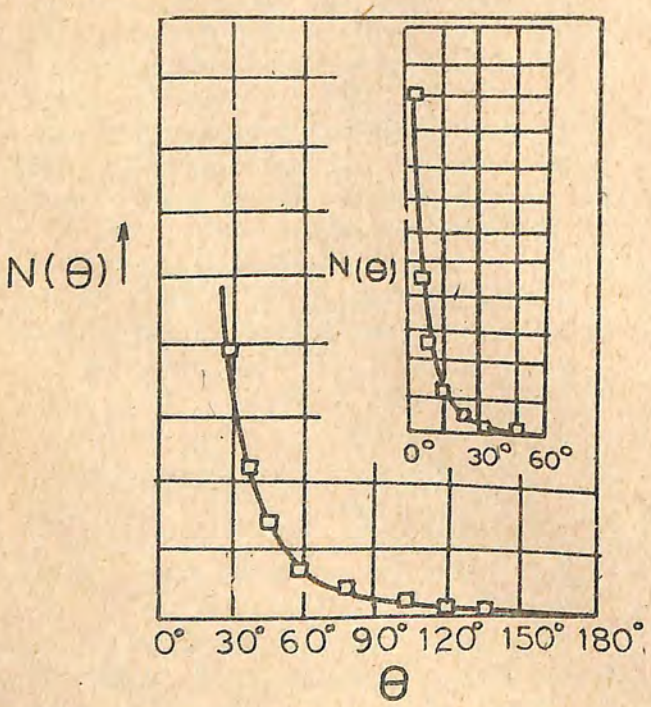


Fig. 9.6 N(θ) versus θ. □ -data of Geiger and Marsden.

the general ideas of Rutherford, which led him to propose a new model—the nucleus model—for an atom. The model is shown in Fig. 9.7. The central circle represents the

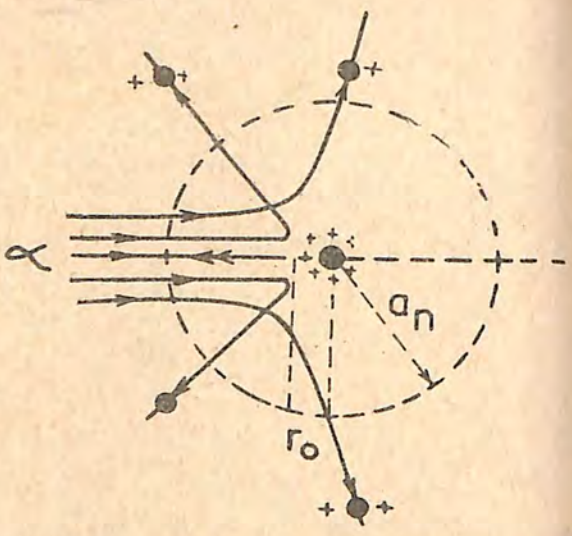


Fig. 9.7 The scattering of α-particles on the Rutherford model.

concentration of a positive charge; this is the nucleus of the atom. The outer circle represents the size of the atom. Scatterings of the α-particles are also indicated in the figure. The figure does not represent the nucleus as well as the atom to the scale. In the above model, practically the whole mass of the atom is in the nucleus and electrons reside outside the nucleus.

To estimate approximately the radius of the nucleus, let us consider an α-particle of energy E MeV incident on the centre of the nucleus. This α-particle will retrace its path after reaching a minimum distance r_o from the centre of the nucleus. Why? Because the kinetic energy E of the α-particle is transformed into potential energy due to coulomb repulsion between the α-particle and the nucleus. At a distance r_o, the distance of the minimum approach to the nucleus, the kinetic energy of the α-particle is equal to the potential energy in the α-particle-nucleus system. Therefore, we have

$$E = \frac{1}{2} mu^2 = \frac{2kZe^2}{r_o} \quad \ldots (9.5)$$

where k is a constant and Z is the atomic number of the foil atoms.

EXAMPLE

Estimate the radius of a gold nucleus when an α-particle of energy 6 MeV is incident on its centre and gets scattered through 180°.

We know, $Z_1 = 2, Z_2 = 79$; $k = 9 \times 10^9$ $Nm^2/coul^2$; $e = 1.60 \times 10^{-19}$ coul; 1 MeV $= 1.6 \times 10^{-13}$ joules

$$r_o = \frac{k Z_1 Z_2 e^2}{E}$$

$$= \frac{9 \times 10^9 \times 2 \times 79 \times (1.60 \times 10^{-19})^2}{6 \times 1.6 \times 10^{-13}} m$$

$$r_o = 6.1 \times 10^{-14} m$$

Thus the radius of the nucleus must be smaller than r_o. The radius of a nucleus can be expressed in fermi (f) units. One *f* is 10^{-15} m.

In conclusion, we have from Rutherford's analysis that (a) the positive charge of an atom is confined to a small volume called *nucleus*, (b) the radius of the nuclei is ~ a few fermi, and (c) electrons reside outside the nucleus.

9.4 Bohr's Theory for the Hydrogen Atom

Bohr, accepting the nucleus model of an atom as well as Planck's quantum theory, proposed an atom model to explain radiations emitted by the hydrogen atoms. His model for the hydrogen atom also called *planetary model of the atom*, represents an electron revolving round a nucleus (a proton in this case) in various circles of a radius a_n. Experimentally, hydrogen atoms emit a series of discrete radiations known as spectral series and hydrogen atoms are stable; these facts are contrary to the classical theory predictions according to which a rotating electron should loose energy and fall into the nucleus Bohr postulated new ideas, which are revolutionary from the classical point of view.

Postulates of the Theory

The postulates are: (i) Of all possible circular orbits allowed by the classical theory, electrons in an atom are permitted to circulate in fixed orbits (stationary orbits) in accordance with the prescribed rules, i e. mvr, the angular momentum is an integral multiple of $h/2\pi$ (ii) electrons in the stationary orbits do not radiate energy, and (iii) an atom may absorb or emit discrete energy photons only when one of the electrons of the atom jumps from the lower to the higher or from the lower orbits, respectively. None of the postulates follows from the classical theory of oscillators or electromagnetism. On the other hand, they violate established laws of the classical theory.

A sketch of the hydrogen atom is shown in Fig. 9.8. Let a_n and v_n be the radius and velocity of the electron in its n^{th} orbit around the proton. The radius a_n is measured from

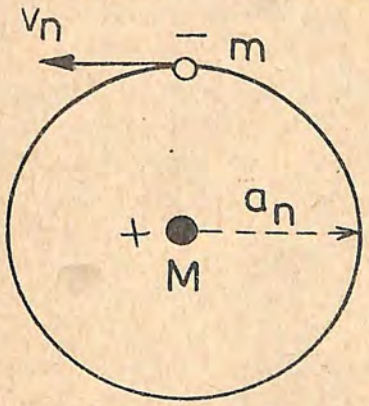


Fig. 9.8 A model of the hydrogen atom.

the centre of the nucleus. For dynamical equilibrium, the centripetal force on the electron should be equal to the electrostatic force of attraction to keep the electron in its orbit. Therefore, we have

$$\frac{kZe^2}{a_n^2} = \frac{m v_n^2}{a_n} \quad \ldots (9.6)$$

where Z is the atomic number of the nucleus for H, Z=1, and *m* is the mass of the electron. Bohr introduced a new rule here that the angular momentum L_n of an electron in

an orbit is an integral multiple of $h/2\pi$, the Planck's constant. That is,

$$L_n = mv_n a_n = \frac{nh}{2\pi} \quad \ldots (9.7)$$

for the hydrogen atom. From the equations (9.6) and (9.7), it follows,

$$v_n = \frac{2\pi kZe^2}{nh} \text{ and}$$

$$a_n = \frac{n^2h^2}{4\pi^2 mkZe^2}. \quad \ldots (9.8)$$

The radius of the electron's lowest orbit in H atom (n=1) is 0.53 A°. The permitted orbits of the first postulate are now given. by the positive integer values of *n*, which are called the *principal quantum numbers of the system.*

The kinetic energy of the electron in its nth orbit is

$$K = \frac{kZe^2}{2a_n}$$

and its potential energy is $\Phi = -kZe^2/a_n$. The total energy of the electron in its nth orbit of the atom is

$$E_n = -\frac{kZe^2}{2a_n}$$

$$= -\frac{2\pi^2 k^2 Z^2 e^4}{n^2h^2} m. \quad \ldots (9.9)$$

Using the third postulate, it can be shown that the frequency ν of radiation from the hydrogen atoms is given by

$$\nu = \frac{2\pi^2 k^2 mZ^2 e^4}{h^3} \left\{\frac{1}{n_1^2} - \frac{1}{n_2^2}\right\} \quad \ldots (9.10)$$

where n_1 and n_2 are the lower and higher principal quantum numbers of energy states of the electron in the hydrogen atom and n_2 is greater than n_1. The energy level diagram for the hydrogen atom is given in Fig. 9.9. The origin of spectral series observed by Balmer and Paschen are indicated in the diagram. The frequencies measured experimentally agree very well with the calculations from the equation 9.10. The energy required to remove an electron from n = 1 to n = ∞ orbit is called *ionisation energy* or *ionisation potential* for the hydrogen atom. Its value for the hydrogen atom is 13.6 eV. A good theory not only accounts for the existing data but also predicts new facts or observations which could be experimentally verified. Below, we mention one of the experimental observations which fitted well in the theory. Brackett (1922) and Pfund (1924) discovered two new series for hydrogen which exactly matched with the calculations given by the equation 9.10. These series were observed only after calculating the expected regions for the series theoretically.

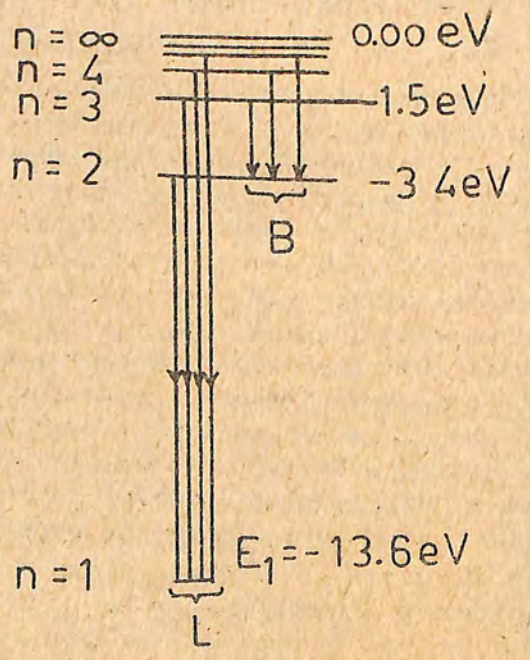


Fig. 9.9 An energy level diagram of the hydrogen atom: B—Balmer series; L—Lyman series.

Limitations of Bohr's Theory

In spite of its success, the theory has limitations. It is not clear why circular orbits have to be chosen when elliptical orbits are also possible. This implies that the theory is not general and complete. Some spectral lines of hydrogen are not single lines but a group of closely packed lines with slightly different frequencies. The theory could not account for such fine features of lines in the hydrogen atom.

It is now known that the planetary model of the atom is not a correct representation; orbits of electrons cannot be exactly defined as is done here. Electrons exhibit wave properties also and the picture of the charge distribution of the electrons in different orbits is drastically different from what is presented here.

With all its drawbacks, this dynamical model was useful to pave the way for further development of the atomic models in modern physics.

9.5 Electron Configurations in Atoms

In an earlier section, we have seen that Bohr's theory for the hydrogen atom satisfactorily explains the discrete radiations emitted by the hydrogen atoms: this theory is a useful guide for understanding the structures and spectral radiations of other atoms.

The atomic radius and the first ionisation potentials were measured for many elements; these quantities as a function of the atomic number (Z) of the elements are shown in Figs. 9.10 and 9.11, respectively. The atomic radius should increase with the atomic number (equation 9.8) and the ionisation potentials of the atoms should approximately vary as Z^2 of the atoms (equation 9.9). Such predictions on Bohr's theory were not experimentally observed. Further, the spectral series of the hydrogen atoms have some resemblances with the spectral series of the alkali atoms. Based on the above facts, Bohr and Stoner concluded that all the electrons in an atom do not occupy the same orbit

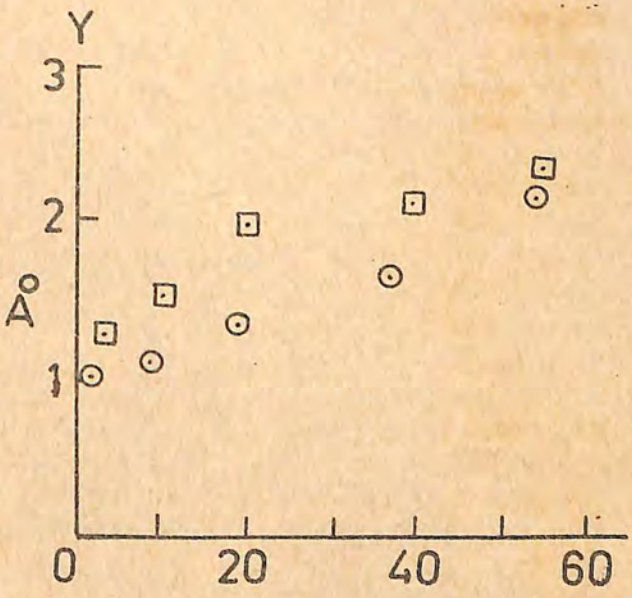


Fig. 9.10 Atomic number versus atomic radius: OX—Atomic number; OY—atomic radius.
⊡ alkali metal ⊙ inert gas.

and the maximum number of electrons in nth orbit of an atom is given by $2n^2$, where n is the principal quantum number. A diagrammatic representation of some atoms, according to Bohr-and-Stoner scheme, is shown in Fig. 9.12: Orbits with n = 1, 2, 3, 4, etc., are called K, L, M and N, etc., shells respectively, and the electrons in orbits with n = 1 can correspondingly be defined as K shell electrons, and so on.

The outermost electrons in an incomplete shell ($< 2n^2$) are called *valence electrons*. Several heavy elements showed departures from the above prescribed rule. Thus, the Bohr-and-Stoner emperical rule is a useful guide for further development of this subject.

Pauli's Exclusion Principle and Quantum Number

By assuming that the orbit of an electron was circular and introducing the principal

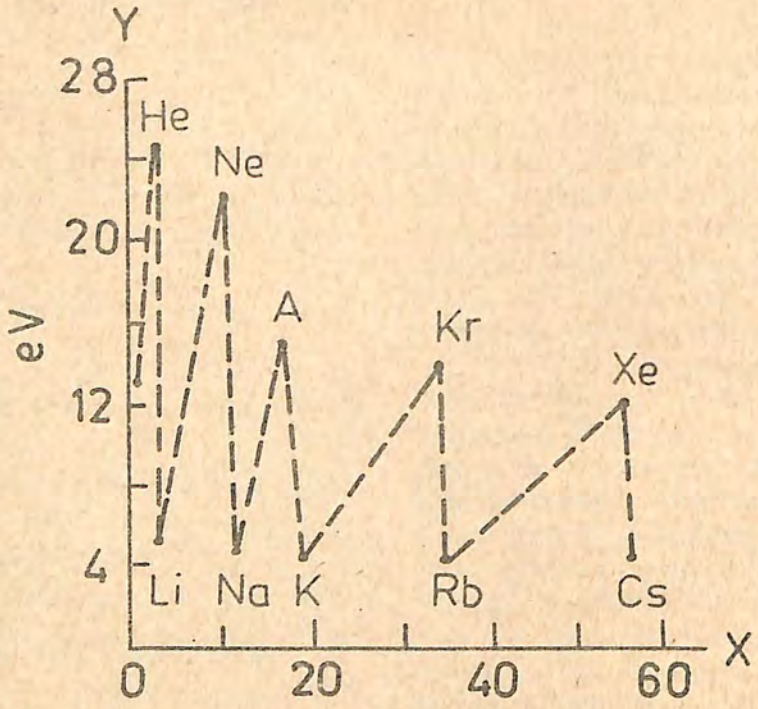


Fig. 9.11 Atomic number versus ionisation potential: OX—atomic number; OY—ionisation potential.

quantum number n, Bohr could arrive at the total energy of an electron in an atom. Sommerfeld extended Bohr's model and solved the general problem in which circular as well as elliptical orbits are permissible for electrons in an atom. Sommerfeld's theory retained the principal quantum number n and introduced a new quantum number k, called the azimuthal quantum number. The ratio n/k was found to be equal to the ratio of the magnitude of the semi-major and semi-minor axis of the corresponding elliptical orbit. It was interpreted to mean that the principal quantum number n was a measure of the major axis of the elliptical orbit, whereas the azimuthal quantum number that of the minor axis.

In Sommerfeld's theory, k could take the values 1, 2, ... n. $k = 0$ was ruled out as it required an electron to traverse the nucleus, since $k = 0$ means zero angular momentum (mvr) and that is possible only if the normal distance r of the electron is zero from the nucleus. The experimental observation on atomic spectra shows that the azimuthal quantum number can have n possible values and can be zero. This serious discrepancy was resolved by the introduction *of the quantum number l*, also called *the azimuthal quantum number* or *the reduced quantum number*. For a given value of n, l takes positive integer values 0, 1, 2, ... $n-1$; that is, $l = k - 1$ or $n - 1$. For example, for $n = 5$, l can assume the values 0, 1, 2, 3, 4.

With the development of high resolution spectroscopy, it was discovered that a Balmer series line, for example, was not single, but consisted of six lines, separate and distinct, but all of very nearly the same frequency. For its explanation, it became necessary to imagine each electron as a sphere spinning about its own axis and attribute a spin half ($s = 1/2$) where s is called the *spin quantum number*.

Spectroscopists like Zeeman discovered

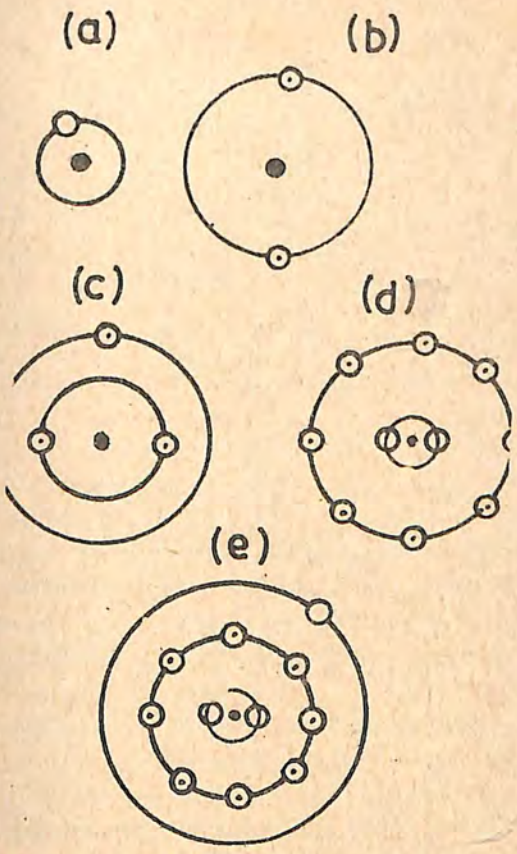


Fig. 9.12 A model of atoms:
(a) Hydrogen (b) Helium (c) Lithium (d) Neon
(e) Sodium. Code: O—Nucleus: O—electron.

that many normally single spectral lines split up into several lines when the gas emitting them was subjected to a strong magnetic field. A way was found to interpret this too, in terms of the Bohr model. An electron in an orbit behaves like a little magnet, the moment being perpendicular to the plane of the orbit. The situation is similar to a current-carrying-coil. In a strong magnetic field, the little magnets try to align themselves along the direction of the field. However, it was found that the magnetic moments (and hence the orientation of the orbit) could not be pointed just in any direction. This, too, is quantised. The electron's *magnetic moment*, is proportional to l. For a given value of l, m_l is found to assume $(2l+1)$ integer values from $-l$ to $+l$, i.e., l, ...0 -1, -2 ... $-l$. For example, for $l = 2$, m_l has five values, -2, -1, 0, 1, 2.

In connection with the fine splitting of spectral lines, the idea of an electron spin was introduced. Like an orbiting electron, a spinning electron is also associated with a magnetic moment. In the presence of a magnetic field, again, it was found that it could take up only discrete orientations. The explanation of the spectral lines required that the vector representing the magnetic moment due to the spin could take up only two orientations, parallel to the field or against it. This gives rise to the fourth quantum number m_s which can be either $+1/2$ or $-1/2$. In conclusion, four quantum numbers, n, l, m_l and m_s are necessary in the theory to completely specify the state of an electron in an atom.

Pauli's exclusion principle states that no two electrons in an atom can have the same four quantum numbers. The configuration of electrons in an atom can be written with the help of the exclusion principle. All electrons with the same n constitute a *shell*. There can be a maximum of $2n^2$ electrons in a shell. Historically, the $n = 1$ shell is known as the K shell, the $n = 2$ shell is the L shell, and so on. The electrons corresponding to $l = 0, 1, 2, 3,$ etc., are, respectively, designated as s, p, d, f. A few configurations for elements are illustrated in Table 9.1.

TABLE 9.1

Shell	K	L		M			N	Configuration
n	1	2		3			4	
l	0	0	1	0	1	2	0, 1, 2, 3,	
Element								
1 H	1							$1s^1$
2 He	2							$1s^2$
3 Li	2	1						$1s^2 2s^1$
10 Ne	2	2	6					$1s^2 2s^2 2p^6$ = Ne
11 Na	2	2	6	1				$Ne 3s^1$
18 Ar	2	2	6	2	6			$Ne 3s^2 3p^6$ = Ar
19 K	2	2	6	2	6			$Ar 4s^1$

The number just behind the symbols, such as *p*, gives the principal quantum number or the shell number and the superscripts, the number of electrons belonging to it.

9.6 X-rays

X-rays were accidentally discovered by Röntgen, in 1895, during his studies on the electrical discharge phenomena in gases. Crystals of platino-cynide, situated near the discharge tube, exhibited bright fluorescence. Even though the discharge tube was covered with black paper to prevent the visible light from the tube, and was located in a dark room, the crystals showed bright fluorescence whenever the discharge tube was in operation; this clearly indicated that some unknown radiation (X-rays) from the tube were responsible for the fluorescence of the crystals. Further studies by Röntgen established that X-rays were generated by the impact of electrons on the walls of the discharge tube.

The applications of X-rays were so important that within weeks of their discovery, they were used as an aid in surgical operations. This single discovery was a good example of the utility of scientific research in daily life.

Production of X-rays

The present X-ray tube does not bear any resemblance to the earlier tubes used by Röntgen and others. A sketch of the modern X-ray tube is shown in Fig. 9.14. The cathode and anode, enclosed inside an evacuated glass vessel, are connected to a high tension of about one lakh volts. The voltage applied across the electrodes is non-oscillatory (direct current voltage). The cathode is given the shape of a concave mirror so as to focus an electron beam on the anode. X-rays are generated from a small spot on the anode and are spread in all possible directions.

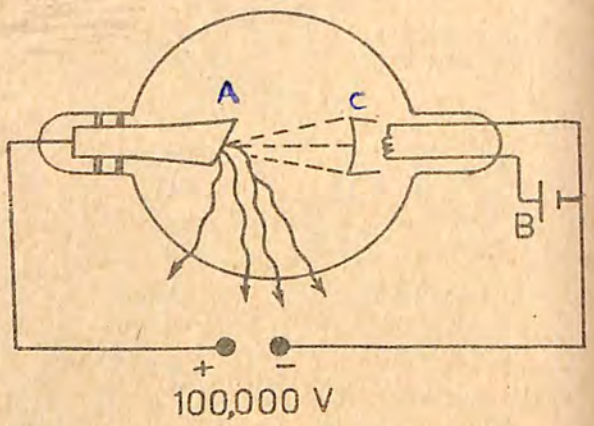


Fig. 9.14 An X-ray tube.

Considerable improvement in the design of X-ray tubes was made by Coolidge in 1913. A filament, heated to incadescence by a battery or a low voltage transformer, is inserted inside the hollow cathode. This arrangement gives rise to an intense electron beam from the cathode. The anode is made of a solid copper bar (Fig. 9.14). The focused electron beam heats the anode considerably and often melts it. To overcome this difficulty, high melting point metals like platinum are embedded in the anode surface. Thus, the electron beam is incident on the platinum target and the heat generated is dissipated by the copper anode; or, sometimes, anode is surrounded by water-cooled coils to avoid excessive heating of the anode.

X-rays are produced by the sudden stopping of the electron beam by the target platinum; the larger the intensity of the electron beam, the greater is the intensity of the X-rays generated. The quality (wave length distribution) of X-rays depends on the direct voltage applied across the tube.

X-ray Spectra

Before we discuss X-ray spectra, it is useful to learn how X-ray wave lengths are measured. X-rays are electromagnetic waves with wave lengths from a fraction of an Angstrom (A°) to about hundred Angstroms. The spacing of atoms or molecules in crystals

is of the order of a few Angstroms. Thus, crystals are used to measure the wavelengths of X-rays, just like gratings are used to determine the wave lengths of the visible radiation.

A simplified diagram of a device—X-ray spectrometer—to determine the wave lengths of X-rays is shown in Fig. 9.15. X-rays from a tube, narrowed down to a fine line by means of stopper slits B and C, are incident on a crystal at an angle θ. The reflected X-rays pass through a detector as indicated in the figure. The positions of the X-ray tube and the detector are fixed and the crystal, mounted on a turn table, can be rotated about its central axis. The relation between d, the distance between the planes passing through the molecules in the crystal, λ the wave lengths of the X-rays and θ, the angle of the reflected X-rays is given by Bragg's rule :

$$2d \sin \theta = n\lambda \qquad \ldots(9.11)$$

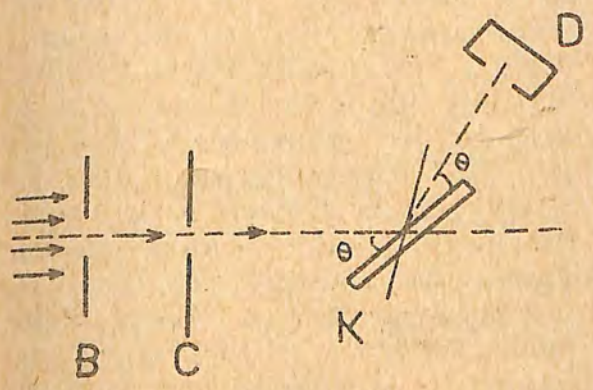


Fig. 9.15 A sketch of an X-ray spectrometer: k—crystal; D—detector.

where n is the order of the spectrum. X-rays, for a given order n, of different wave lengths get reflected through different angles θ. From the known values of d, and the measured values of θ on an X-ray spectrometer, the values of λ are determined from the equation 9.11.

Experimental observations show that X-rays from the X-ray tubes are of two types—characteristic X-ravs and continuous X-rays.

Characteristic X-rays

The spectra of this group comprise of several radiations with the specific sharp wave lengths similar to the spectral lines of atoms like hydrogen. The wave lengths of this group represent characteristic discrete radiations emitted by the atoms of the anode. Thus, characteristic X-rays are utilized to identify the atoms of the anode; this identification is similar to the identification of the atoms from the visible radiations emitted by them.

Origin of Characteristic X-rays: The equation 9.10 of section 9.3 represents the frequencies of radiations emitted by atoms in terms of constants and their principal quantum numbers. The expression can be written in the following simpler form:

$$\nu = RZ_e^2 \left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

where R is constant and Z_e is the effective atomic number of a given element.*

Consider a nucleus of the atomic number Z; electrons around the nucleus are distributed in the prescribed orbits according to the Pauli Exclusion Principle. If an incident X-ray collides with an electron in the K shell ($n = 1$) of the atom and removes it to a large distance (ionisation of the atom where an electron from the K shell is knocked out), then the K shell is short of one electron. Another electron from L, M or N shell can make a transition to the K shell to complete the K shell ($2n^2$): the incomplete L, M or N shell, in turn gets an electron from still higher shells. This process of the jumps of an electron from higher complete shells to lower incomplete shells proceeds till inner shells are filled up. Thus, radiations are generated as a result of several transitions. Some of the radiations emitted happen to be in the X-ray wave length region. Transitions from $n = 2, 3, 4$ etc., to $n = 1$ shell give rise to X-rays called K_α, K_β, K_γ, etc.,

* Z_e, the effective atomic number is slightly less than the actual atomic number Z of the element.

radiations, respectively. Similarly, transitions from $n = 3, 4, 5$ etc., to $n = 2$ shell generate L_α, L_β, L_γ, etc., X-rays, respectively. These transitions are illustrated in Fig. 9.16. Thus, the origin of characteristic X-rays from atoms have similarities in transitions to Lyman,

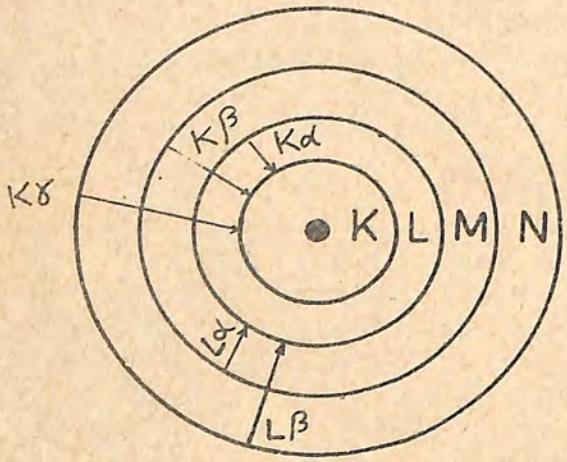


Fig. 9.16 The origin of the characteristic X-rays. diagram is not to scale.

Balmer, etc., series of the hydrogen atoms. Characteristic X-rays are discrete radiations which have distinct wave lengths for each element of the periodic table. Thus, the equation 9.10, derived for the hydrogen atom, explains the origin of characteristic X-rays from the elements of large Z values. If we plot frequencies of K_α or K_β radiations, from different elements, as a function of a square of the atomic number of the respective elements, we obtain a straight line relation: $\nu = RZ^2$. This expression is the famous Moseley's law in X-rays. The value of the constant R is the same as that in the equation 9.10.

Continuous X-rays

In addition, the X-ray tubes emit X-rays of all wave lengths, which appear as background radiations over which characteristic X-rays are superimposed. This background or continuous radiations are called continuous X-rays. One important feature of continuous X-rays is that they end abruptly at a certain lower wave length; this abrupt cut off appears as a sharp edge for the background X-rays on the photographic film. The wave length (λ_c) for the sharp edge depends only on the voltage applied across the tube.

Origin of Continuous X-rays: If an electron beam of energy eV electron volts is incident on the target atoms of the anode, it interacts with the coulomb field of electrons in the atomic orbits as well as with the coulomb field of the nucleus. Since the nucleus has a concentrated positive charge Ze on it, its interaction with the incident electrons is large and the force in the interaction is Ze/r^2, where r is the distance between the nucleus and an electron of the incident beam. This force results in the acceleration of the electron given by $Ze/r^2 = ma$, where m and a are the mass and acceleration, respectively, of the electron. The classical theory of electromagnetism states that the suddenly accelerated electrons emit electromagnetic radiations (this process is referred to as Bremsstrahlung process) and the wave lengths of radiations are continuous. Thus, the energy spectrum of X-rays by the Bremsstrahlung process is continuous and the maximum frequency of X radiation $h\nu = eV$ is obtained only when the energy eV of the incident electron is given to a single X-ray photon. The experimental results of continuous X-rays (intensity *versus* wave length) are shown in Fig. 9.17. The experimental results are in agreement with the classical theory outlined above.

Properties of X-rays

The technological and research utility of X-rays lies in their ability to (a) ionise atoms and molecules of matter, (b) cause fluorescence in certain chemical compounds,

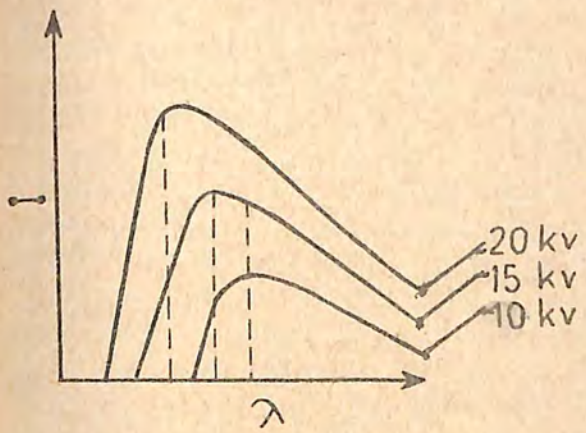


Fig. 9.17 A spectrum of the continuous X-rays. I—intensity, λ—wavelength. **Cutoff at short λ ends is given by hν=eV.**

(c) effect photographic films, and (d) penetrate through solids. Some of these properties are described here with a view to making clear the application aspects of X-rays given in the later section.

A charged electroscope is discharged when X-rays are incident on it. The explanation for this observation is that X-rays ionise atoms and molecules of the air medium in the electroscope. The ionisations of atoms and molecules result in the formation of positive and negative ions in air. These ions when finally collected on the gold leaves of the electroscope discharge it. This process is utilized not only to detect X-rays but also to measure the intensity of X-rays. The measurement is possible because the total ionization is proportional to the intensity.

X-rays, when incident on special photographic films (X-ray films), affect silver halide crystals, suspended in a gelatine medium. The affected crystals, after developing the film, appear as black specks. Thus, the images cast by X-rays can be recorded on the films. The opacity of the images on the film gives a measure of the intensity of X-rays.

The images due to the X-rays passing through an object are made visible on a screen coated with chemical compounds like zinc sulphide. This is possible because the X-rays incident on the screen cause fluorescence in certain chemical compounds like zinc sulphide. This fluorescence radiation happens to be in the visible light region. The intensity of X-rays is defined as the amount of energy carried per unit time across per unit area, the area being perpendicular to the flow of energy.

The intensity of X-rays decreases with increase in the amount of matter through which it passes. It has been experimentally found that the elements of higher atomic numbers absorb X-rays to a very much greater extent than the elements with low atomic numbers. X-rays with large penetrating capacity are called *hard X-rays* while those with small penetrating power are called *soft X-rays*. The absorption of X-rays is mostly due to ionisation, excitation, fluorescence, etc., of atoms, molecules, crystals, etc., of the absorber.

Applications of X-rays

X-rays have important and useful applications in surgery, medicine, engineering, and studies of crystal structures in particular and solid state studies in general. We describe a few applications here.

X-rays are absorbed more in heavy than in light elements. When X-rays traverse an object like the palm of a hand, the bones containing calcium and phosphorus absorb more than the surrounding tissues which comprise of light elements like hydrogen, carbon, oxygen. Thus, the image of a hand on a photographic film or fluorescent screen vividly brings out the details of the bone structure of the hand. Such an image allows detection of cracks, fractures of bones, dislocation of bone joints. Thus, surgeons correct the defects after studying the photographs of any part of a human body.

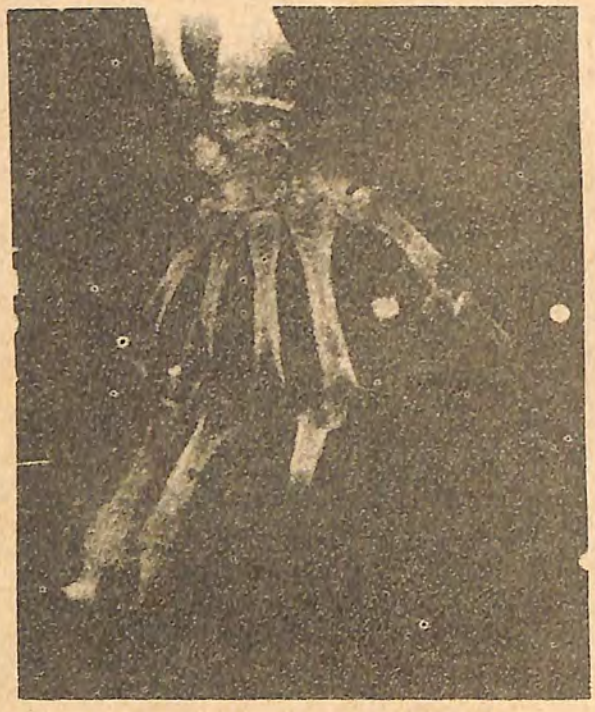


Fig. 9.18 An X-ray photograph of a hand.

X-rays cause damage to the tissues of a body, for molecular chains and cells are ionised and broken by X-rays. Certain malignant growths like cancer in a human body is dangerous to life. Such growths can also be damaged by X-rays. Thus, by proper X-ray therapy, sometimes, the growth of cancerous tissues can be checked and the disease can be cured if detected in early stages.

Earlier, we saw how X-ray wave lengths can be measured by an X-ray spectrometer. We can reverse the method and study the spacing of atoms and molecules in a crystal. Thus, using monochromatic X-rays, the structures of crystals (arrangement of atoms and molecules in crystals) are studied. By a similar procedure, the structures of solids like glass, diamond, fibre, metals, etc., are also examined. Defects in metals and alloys can also be studied by X-rays, for the defective regions have a different structure formation than normal metals and alloys.

9.7 Photo Electric Effect

Substances when exposed to electromagnetic radiation such as X-rays, γ-rays, ultra-violet visible and infra-red light, absorb the radiation, and electrons are emitted from the surface of the substance. These phenomena are called *photo-electric effect* and the emitted electrons are referred to as *photoelectrons*. This effect is one of the general phenomena which occurs when electromagnetic radiation interacts with matter. Different substances emit photoelectrons only when exposed to different wave length regions of the radiation. For example, X-rays eject photoelectrons from K and L shells of heavy elements; photoeletrons are emitted when visible and ultraviolet light is incident on alkali metals. In this section, we confine our attention to the study of this effect in metals using ultraviolet and visible light.

Experimental Studies

A simple device to study the photoelectric emission is shown in Fig 9.19. Two zinc plates C and D. enclosed in a quartz vessel F, are connected to a battery B and a micrometer A as indicated in the figure. The vessel with the electrodes system is called a photo tube. If light is shown on the cathode D, the current I flows but the light incident on the anode B does not produce a current in the circuit. Thus, it is evident that the negatively charged plate loses its charge, when the light is incident on it and the current in the circuit is due to the flow of electrons across the electrodes. I is a measure of the number of electrons (I/e) collected at the positive electrode C.

The dependence of the number and energy of the photoelectrons from a metal on the intensity and frequency of the radiation respectively constitutes a detailed study of this effect.

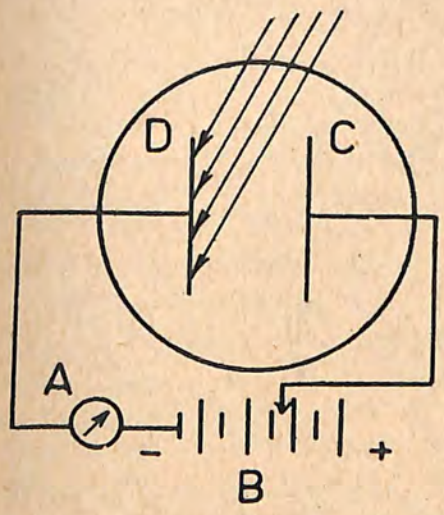


Fig. 9.19 A sketch of a photo-tube to study the photoelectric effect.

It is observed experimentally that for a given spectral quality of light, the photo-electric current I in microamperes is directly proportional to the intensity of light. Let V

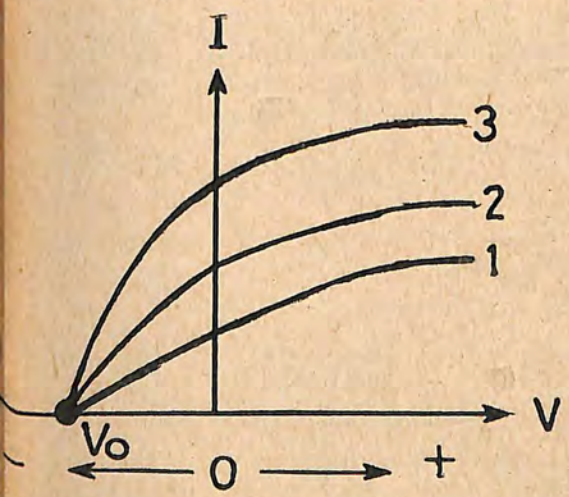


Fig. 9.20 Photo current versus voltage between the electrodes of a photo-tube.

be the potential on C relative to the electrode and V be the given small values, positive as well as negative. The variation of the photo-current as a function of the voltage and intensity of radiation is shown in Fig. 9.20. Curves 1, 2, and 3 in the figure are for increasing values of intensity. We observe that (i) the photocurrent I saturates above a certain value of V; the value of the saturation current is proportional to the incident light intensity and (ii) for the negative value of V, the photocurrent decreases and finally becomes zero at a negative voltage V_o. V_o is called the retarding potential or cut-off potential. Below this value for V_o no photocurrent is observed, irrespective of the radiation.

The retarding potential V_o depends on the nature of the metal as well as on the frequency ν of radiation. For a given frequency ν, V_o measures the maximum kinetic energy K_{max} of the photoelectrons that just reach the anode C. This is due to the fact that the electrons have to move in a potential, which opposes their motion. Therefore, we have,

$$E = K_{max} = \frac{1}{2} mu_{max}^2 = eV_o \ldots (9.14)$$

where e, m and u_{max} are, respectively, the charge, the mass and the maximum speed of the electrons. K_{max} can be expressed in joules per coulomb (V_o/e). The retarding potential V_o/e as a function of the frequency of radiation (ν) is measured and the results are shown in Fig. 9.21. The results for three different metals are given by the lines 1, 2 and 3 in the figure. For each metal, there is a certain frequency ν_i called the threshold frequency at which the photocurrent in the tube is zero. In this figure, ν_1, ν_2 and ν_3 ($\nu_3 > \nu_2 > \nu_1$) are the threshold frequencies for the three metals. An important feature of Fig. 9.21 is that the slopes of the lines, are the same and are equal to h, but the intercepts of the lines on the E-axis are different for different metals. The equation for the lines can be written as

$$E_i = h\nu C_i \qquad \ldots (9.15)$$

where E_i is the energy of the photoelectrons

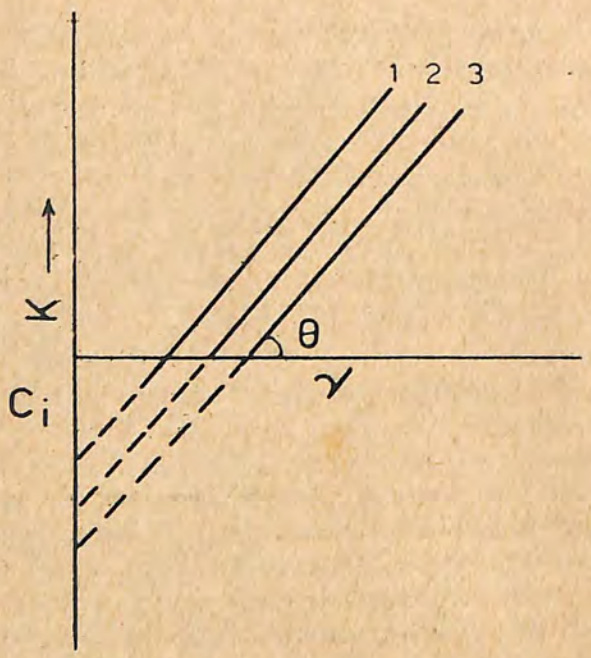


Fig. 9.21 Frequency versus retarding potential in units Vo/e.

and C_i is the intercept on the E-axis for the metal *i*.

A summary of the above results, which any acceptable theory on photoelectric effect must explain, are listed here. They are: (i) For a given metal and frequency of light, the number of photoelectrons is directly proportional to the intensity of light. (ii) For every substance, there is a definite threshold frequency (ν_i) below which no photoelectrons are emitted, irrespective of the intensity of light. (iii) Above the threshold frequency, the energy of the electrons is proportional to the frequency of light.

Einstein's Theory for Photoelectric Effect

Before the advent of the quantum theory, attempts were made to interpret the above results on the basis of the classical theory of electromagnetism. According to the theory, the intensity of the electromagnetic waves (light) is a function of the amplitude of the waves. The wave concept of light accounted for various phenomena, like the interference and the diffraction of light. If the emission of photoelectrons is due to the interaction between the electromagnetic waves and the electrons of a metal, the number and energy of the photoelectrons should only depend on the intensity of light, which is contrary to experimental results. Thus, the classical theory could not account for observations on photoelectric effect.

In 1900, Planck postulated that energy E of a harmonic oscillator of frequency ν is given by the expression, $E = h\nu$ where h is the Planck's constant. Each light wave with the energy E is called a *photon*. According to this picture, the intensity of monochromatic light of a frequency ν_1 is regarded as the number of photons, each with the fixed energy E_1.

Einstein (1905) made use of Planck's quantum theory to interpret the photoelectric effect. His theory is as follows. For each photon of light absorbed in metal, a photoelectron is emitted: Energy has to be expended to remove an electron from inside to outside the metal surface. Thus, the kinetic energy K of the electron is given by

$$K = h\nu - W \qquad \ldots (9.16)$$

This equation is also called Einstein's equation. W is the work function of the metal and it represents the energy loss for the photoelectrons in emerging out of the metal surface. Usually, the incident light penetrates a distance of a few hundred atomic layers (10^{-8} m) and the photoelectrons emitted from different depths will undergo collisions, with the result that the photoelectrons have energy distribution outside the metal surface. This fact accounts for the variation of I with the potential V (Fig. 9.20). The minimum energy or threshold frequency which a photon must have to remove an electron (K = 0)

from the metal should be equal to the work function W of the metal (i.e., $h\nu = W$ from the equation 9.16). Below the threshold frequency ($\nu < W/h$), no photoelectron is emitted for any intensity of light. Above the threshold, the kinetic energy K of the electrons is linearly dependent on the frequency of light and the number of photoelectrons is proportional to the intensity of light. Thus, all experimental results are accounted for in Einstein's theory.

The work function for some metals is listed in the following table.

TABLE 9.2

Metal	Na	K	Ce	Zn	Fe	Ni
W (eV)	2.5	2.3	1.8	3.4	4.8	5.9

Applications of Phototubes

A phototube converts light energy into electrical energy. It is a must in television transmitting stations. Light reflected from objects or drama scenes is focused on a suitable phototube. The electrical energy is again converted into a suitable form of electromagnetic waves for transmission.

Cinema films involve two synchronized processes, namely, the picture and the sound. Photographic scenes and action pictures on films is well known but the recording of sound in synchronization with action is a must for film industry. Sound waves are converted into electrical waves and electrical waves into light waves. These light waves or signals are photographed on the film along with the action picture. By a reverse process, we get the picture and the synchronized sound in a movie theatre.

An interrupted light beam (invisible, of course) across a door can be made to sound a burglar alarm by means of photocells.

A photo-tube has advantage over eye (i) it gives qualitative results and (ii) photo currents can be amplified. Thus photocells are advantageous in the study of astronomical phenomena like temperature and spectra of stars. It can be used to control temperatures of furnaces and chemical reactions. It has advantages in automatic control and check systems like street lighting, traffic signals, speeds of cars, etc.

9.8 Dual Nature of Radiation and Matter

Dual Nature of Radiation

It is established that electromagnetic radiation behaves as though it is composed of waves. Interference, diffraction, etc., are effects that are the direct consequences of the wave properties of radiation. In photoelectric effect, a single photon (wave radiation) gives up all its energy to a single electron and this energy partially appears as the kinetic energy of an electron. It appears as if a particle (photon) imparted energy to another in a collision. Thus, in photoelectric effect, radiation exhibits particle properties. It is puzzling that radiation has particle and wave attributes, which are not reconcilable. The basic question is whether radiation is a wave or a particle or both.

de Broglie Waves (1923)

The ultimate structural components of matter such as electrons, protons and neutrons behave like particles. The fundamental forms of the physical universe are mass and energy and they are intimately related (Einstein's mass energy principle). Thus, there must be a mutual symmetry between matter and energy. The radiant energy has dual aspects of waves and particles and, therefore, from symmetry considerations de Broglie predicted that matter must also have dual nature. Matter in motion must be accompanied by waves called the *de Broglie waves*. From various considerations, he

derived a relation between the wave length λ of matter waves and the momentum (mv) of matter. The relation is

$$\lambda = \frac{h}{mv} \qquad \ldots (9.17)$$

where h is Planck's constant.

Experimental Verification of the de Broglie Waves

The tests for interference and diffraction effects in a beam of electrons are similar to those employed for light and X-ray beams. To detect de Broglie waves, we must have an idea of the waves associated with a beam of electrons of a velocity v. The following solved example serves the purpose.

EXAMPLE

Calculate the wavelength of de Broglie waves associated with a beam of electrons, which traverses a potential difference of 300 volts. Assume, the initial speed of electrons is zero.

$$v = \sqrt{\frac{2Ve}{m}}$$

$$= \left\{\frac{2 \times 300 \times 1.60 \times 10^{-19}}{9.11 \times 10^{-31}}\right\}^{1/2}$$

$$v = 1.07 \times 10^{7}\ m/S$$

$$= \frac{h}{mv}$$

$$= \frac{6.63 \times 10^{-34}}{9.11 \times 10^{-31} \times 1.07 \times 10^{7}}$$

$$= .57 \times 10^{-10} m \text{ or } 0.57\ A^{\circ}.$$

To measure the wave lengths of this order, optical gratings are of no use but crystals are ideally suited (refer Section 9.6 on X-rays). Below we describe some experimental methods to test the reality of de Broglie waves.

Davisson and Germer Experiment (1927)

The experimental arrangement, used by Davisson and Germer, is sketched in Fig. 9.22. Electrons of energy ~ 50 eV, from an electron gun, are incident perpendicular on a plane in a nickel chloride crystal. The diffracted beam is collected in a detector and measured. The intensity of the electrons is measured as a function of θ as indicated in the figure. The intensity is maximum for certain values of θ given by the equation

$$n\lambda = 2d \sin\theta$$

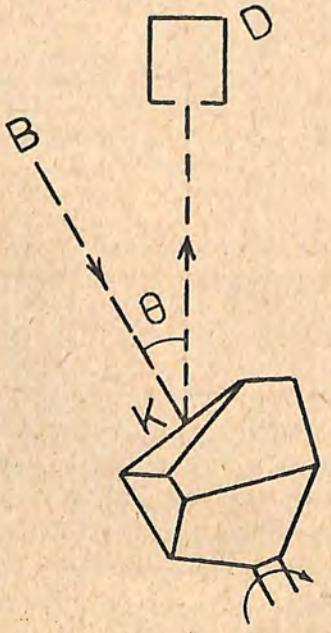


Fig. 9.22 The Davisson and Germer experiment to study the wave nature of electrons. B—electron beam: k—crystal; D—detector.

where n is the order of diffraction and d the atom spacing in crystals. The wave length (1.75 A°) of the electrons is calculated from the de Broglie relation and for a given order (n) of the spectrum and atom spacing (2.5 A°) in the crystal, the expected angles for diffraction of the electron maximum are calculated. In the first order spectrum, the measured values of θ agreed with the calculated values, thereby proving the existence of the de Broglie waves for electrons in motion.

Thomson Experiment

The device used in this experiment is shown in Fig. 9.23. Electrons are incident on a platinum foil. A photographic plate is placed behind the foil as indicated in the figure. The platinum foil is composed of single crystals oriented randomly. Thus, the diffraction pattern of the electrons appears as concentric circles, where the density of the electrons is maximum. The pattern is similar to that of the amorphous solids using

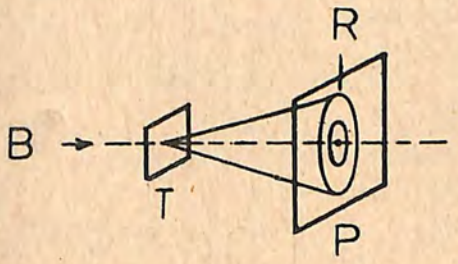


Fig. 9.23 Thompson's diffraction of electron beam experiment. B—electron beam; T—thin foil; p—photographic plate; R—diffraction rings.

X-rays (see Fig. 9.24). The experiment also proved the existence of the de Broglie waves and the correctness of the de Broglie relation.

Later, Easterman and Stern performed similar experiments, using helium atoms, hydrogen molecules, instead of electrons. The reslts were similar to those of the above two experiments. We now know that all particles in motion have wave characteristics in accordance with the de Broglie relation. Thus, the dual nature of matter also is firmly established.

Applications of the de Broglie Waves

For electrons that pass through a potential difference of V volts, the de Broglie wave length λ in A° is given by the expression

$$\lambda = \frac{h}{\sqrt{2me}} \cdot \frac{1}{\sqrt{V}}$$

where m, e and h have the usual meanings. Substituting the values of constants, we get, $\lambda = \frac{12.3}{\sqrt{V}}$ A°. For high energy electrons, λ can be made as small as desired. The resolving power of an optical instrument depends on the wave length of light. Optical microscopes have magnification of ~ 1500 and the resolution of a micrometer. The above figures are mainly restricted due to the wave length of the light used. The de Broglie waves of small wave lengths can be utilized

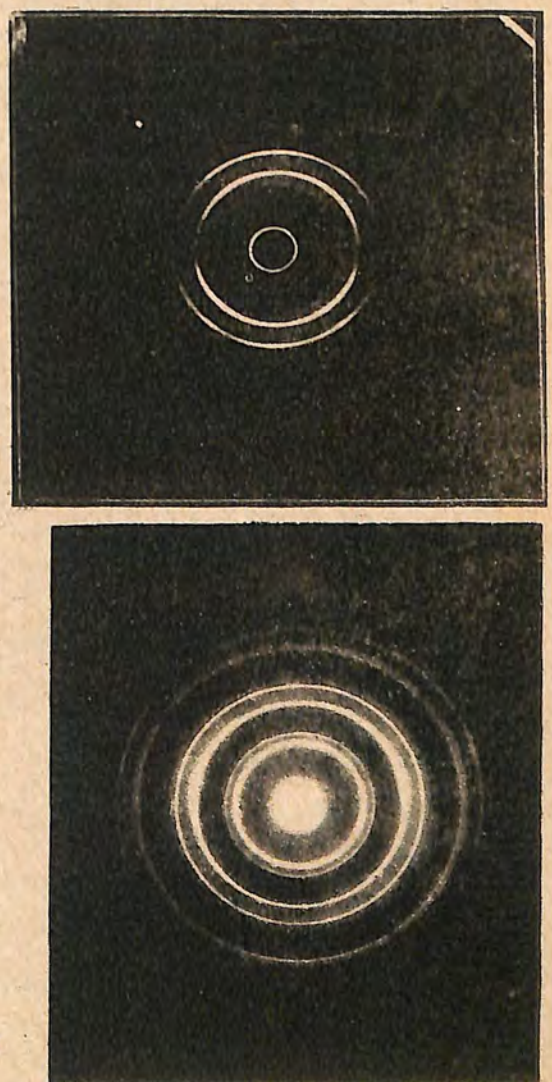


Fig. 9.24 (a) An X-ray photograph of copper wire (transmission). (b) Electron diffraction pattern of copper wire (transmission).

to improve the magnification and resolution of very minute objects.

An instrument using an electron beam, to study very minute objects like viruses, microbes and the crystal structures of solids is known as electron microscope. An electron beam ($\lambda = 0.2$ A°) is focused on an object and its image is brought on a photographic plate. The effects of the appropriately adjusted electric and magnetic fields on the electron beams are similar to those of the lenses on light. An operational sketch of an electron microscope together with that of an optical microscope is shown in Fig. 9.25. An electron microscope has magnification of ~ 100,000. In physics, it is used to study crystal structures and, in biology, to study microbes and viruses.

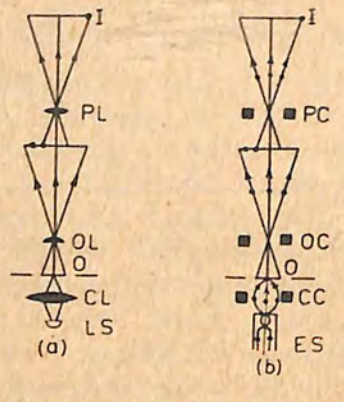


Fig. 9.25 Image formation in an optical and electron microscope.
(a) I—final image; PL—projector lens; OL—objective lens; O—object; CL—condensor lens; LS—light source.
(b) PC—projector coil; OC—objective coil; CC—condensor coil; ES—electron source.

EXERCISES

9.1 The voltage across the electrodes of a cathode ray gun (Fig. 9.2a) is 500 volts. Calculate (i) the energy gained by the electrons, (ii) the speed of the electrons, and (iii) the momentum of the electrons.

(8×10^{-17} J 4/3 $\times$ 10^{7} m/s $1{\cdot}2 \times 10^{-23}$ kg m/s).

9.2 An electron beam of the problem 9.1 passes through a parallel plate condenser (Fig. 9.2b). The electric intensity between the plates is 20 volts/cm, the distance between the plates is 5 cm and the length is 10 cm. Calculate the deflection angle of the beam.

(10^{-3} radian)

9.3 Most mass of the atoms is with the positive charge. In the case of the hyrogen atoms, what fraction of the atomic mass is with the positive charge?

(1833)

9.4 In a head-on collision between an α-particle and a gold nucleus, the minimum distance of approach is 4×10^{-14} m. Calculate the energy of the α-particle.

(5.7 MeV)

9.5 Suppose that the positive charge of an atom, instead of being concentrated in the nucleus, is uniformly distributed over the whole volume of the atom; this is the Thomson model for the atom. Discuss qualitatively α-particle scattering in the Thomson and Rutherford electron in $n = 1$ orbit?

9.6 The classical electromagnetic theory makes some predictions for the hydrogen atoms. List the predictions and compare them with the experimental observations.

9.7 Calculate the radius of a hydrogen atom in its stable state. What is the velocity of an electron in $n = 1$ orbit?

($a_1 = 0.53$ A°. $v_1 = 2.2 \times 10^7$ m/s)

9.8 Calculate the value of the constant $R = 2\pi^2 k^2 me^4/h^3$ in the equation 9.10. What are the dimensions of this constant?

(3.289×10^{15} Hz)

9.9 Show that the ionisation potential for a hydrogen atom is 13.6 eV.

9.10 Calculate the ionisation potential for a lithium atom (see Fig. 9.12). [Hint: $n = 2$ to $n = \infty$ transmission and $Z = 3$.]

(4.4 eV)

9.11 Write the electronic configurations for beryllium, carbon and oxygen atoms.

9.12 For an electron in N shell, list all possible values for the quantum states.

9.13 Calculate the wavelength of a K_α X-ray for gold atoms. (Hint: use the equation 9.10.)

(1.522×10^{19} Hz)

9.14 The angle of reflection for monochromatic X-rays from a crystal whose atomic spacing is 2.5 A° is 15°. Calculate the wave length of the X-rays.

(1.294 A°)

9.15 Calculate the maximum wave length of the continuous X-rays from an X-ray tube whose operating voltage is 50,000 volts.

($\nu = 12.1 \times 10^{18}$ Hz)

9.16 Monochromatic X-rays ($\lambda = 1$ A°) are reduced to 1/3 of their original intensity in passing through a gold foil ($Z = 79$) of 3 mm thickness. Calculate the absorption coefficient for the X-rays. What are the dimensions of the absorption coefficient.

($3.7\ cm^{-1}$)

9.17 The wavelength of a K_α line is 1.54 A° for copper. Calculate the ionisation potential of a K shell electron in copper.

(12.9×10^{-16} J)

9.18 Calculate the energy of a photon whose (i) frequency is 1000 kilo cycles/s (radio waves), (ii) wave length is 6000 A° (yellow light) and (iii) wave length is 0.6 A° (X-rays).

[(i) 6.6×10^{-28} J, (ii) 3.3×10^{-19} J, (iii) 3.3×10^{-15} J]

9.19 What is the frequency of a photon whose energy is 75 eV?

(18×10^{15} Hz)

9.20 Calculate the threshold frequency of photons which can remove photoelectrons from (i) cesium, and (ii) nickel surfaces.

[(i) $\nu_{Cl} = 4 \times 10^{14}$ Hz; (ii) $\nu_{Ni} = 1.6 \times 10^{15}$ Hz]

9.21 If the speed of a photoelectron is 10^4 m/s, what should be the frequency of the incident radiation on a potassium metal?

$(1.2 \times 10^{15}$ Hz)

9.22 The wave length of a photon is 1.4 A°. It collides with an electron. Its wave length after the collision is 2.0 A°. Calculate the energy of the scattered electron.

9.23 Calculate the de Broglie wave lengths for electrons and protons, if their speed is 10^5 m/s.

$[\lambda_e = 7.25 \times 10^{-7}$ m; $\lambda_p = 3.9 \times 10^{-12}$ m]

9.24 What is the momentum of electrons, if their wave length is 2 A°

$(3.3 \times 10^{-24}$ kg m/s)

9.25 The speed of electrons in an electron microscope is 10^5 m/s. If protons with the same speed are used instead of electrons, what additional advantage such a proton microscope has over an electron microscope? Discuss.

UNIT 10

Concepts of Space, Time, and Mass in Relativity

From observations and studies on physical phenomena, like planetary motion, general principles and laws governing them are inferred. The measurable quantities, like energy, momentum, etc., for the phenomena, are derived from three basic quantities, namely, length, time and mass. Certain basic assumptions, which appear most reasonable from the commonsense point of view, are used without question to derive physical laws of nature. These are called implicit assumptions about the nature of space, time and mass. For example, in Newton's mechanics, space intervals, time intervals and the mass of a body do not depend on observers, if their relative motion is uniform. In other words, if there is a clock in a train in uniform motion, the values of its mass, diameter and the interval of time between ticks as measured by an observer in the train and another observer on the ground will be the same. These notions about space, time and mass as conceived by Galileo and Newton had been successful to account satisfactorily for observations in mechanics and astronomy. The same notions and laws when applied to the speed of light did not yield consistent results. Einstein had to re-examine and modify concepts of space, time and mass in order that the laws of nature yielded consistent results in all situations. In this unit, we will present the sequences of change in the notions of space, time and mass from Galileo's time to the present and discuss their impact on physical observations and laws.

10.1 Definition of Observer, Event and Frame of Reference

Before we embark on the main theme of this Unit, it is necessary to understand certain terms which often occur in the subject. An event is a simple or complicated occurrence taking place in space at a given time. For example, the tickings of a clock represent a series of events. Lightening in the sky, collision of two cars on the road, fall of fruits from trees, etc., are some other examples of events that occur in space. By an observer, we mean a person or his measuring equipment to observe or measure events. He is in possession of all gadgets, like scales, clocks, telescopes etc., necessary to perform measurements on events; he draws inferences from observations.

'A fruit fell from a tree' is a qualitative observation of an observer, who saw the fruit falling. If precise information like where the fruit was before it fell, at what time it started to fall, when did it hit the earth, etc., is desired, then the observer will have to make quantitative measurements, for which he must set up a coordinate system or a frame of reference. The observer is located in the frame of reference itself. The observer regards that his frame of reference has no motion, i.e., it is in a state of rest, and other frames of reference are in relative motion. A frame of reference or coordinate system is said to be inertial, when bodies in it follow Newton's law of inertia and other laws of Newtonian

mechanics. Any other frame in rectilinear and uniform motion relative to the inertial frame is also an inertial frame. We often refer it as an inertial system and an observer in it is referred to as an inertial observer.

Normally, for convenience we use a cartesian coordinate system, with an observer at the origin of the system. The coordinates of a point P in space are shown in Fig. 10.1. If the point P is in motion, then the dependence of (x, y, z) on t gives a precise statement of its motion.

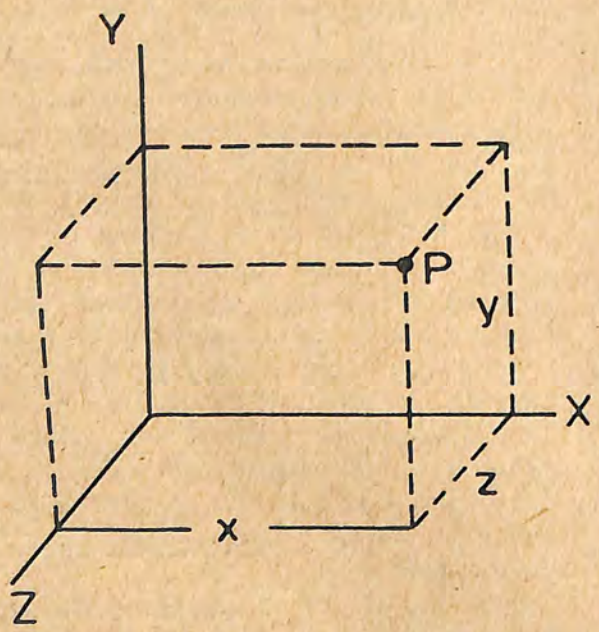


Fig. 10.1 Coordinates of a point P in a cartesian coordinate system.

10.2 The Principle of Relative Motion

The subject-matter of this section is mostly argumentative and the conclusions are based on the generalisations of facts derived from illustrative examples. Let us consider the following examples: (i) Two trains, on the side by side tracks, are at rest relative to a railway platform. An observer is on one train and the other train starts to leave the platform. The observer looks at the other train. Without any additional information, would the observer be able to answer whether his train is in motion or not? (ii) The observer is on an inertial train. The observer looks at the scenery and finds that the trees and telegraph poles are flying past by him. Could the observer state whether his train is moving or not? In the examples (i) and (ii), the observer is not sure whether his train is in motion or not. If the observer in the example (i) looks at the platform, he will find that he is at rest relative to the station and, therefore, he can now conclude definitely that his train is not moving. The observer in the example (ii) is undecided unless he is told that the telegraphic poles and trees are in the state of rest relative to the railway track.

Thus, from the above examples, we arrive at the conclusion that the state of rest or of uniform rectilinear motion (i.e., inertial state) are relative terms and can only be defined relative to the objects outside the inertial frames.

To examine the situation further, we may perform additional experiments, like pouring tea into a cup from a kettle, throwing a ball upwards and catching it on its descent, the motion of a simple pendulum, and firing a bullet shot from a gun and studying its trajectory, etc. The first two experiments are common experience to all of us. It is as easy to pour tea into a cup from a kettle or throw a ball upwards and catch it on its descent in a train at rest as in a train with uniform rectilinear motion. Other experiments conducted inside a train did not reveal any results which could distinguish between the state of rest and of uniform rectilinear motion. Such experiments show that all inertial frames are equivalent in describing the laws of motion, or the laws that govern motion have the same form for all inertial observers. The above statement is called the *principle of relative motion* or *Newton's Relativity Principle*. The mathematical treatment of this principle based on physical reasoning is given in the next two sections.

10.3 Galilean Transformations

In Section 10.1, we indicated how an observer in a frame of reference records the position and time coordinates of a point in space as (x, y, z) and t, respectively. The coordinates of the same point could also be recorded by an observer in relative motion. If we consider the free fall of a body in space, it would appear to move vertically down for an observer at rest on the earth but it would track a parabolic path for an observer on a train with relative uniform rectilinear motion. It is important to study how the two observers in relative motion record events that occur in space; further, it is also relevant to see what conclusions they would make on laws governing events. For this purpose, the need of the coordinate systems is obvious and here below we describe two systems which are often made use of in relativity theories.

Let us consider two inertial observers, O and O′, located at the origins of the inertial frames S and S′, respectively. A diagramatic representation of the inertial frames and observers is given in Fig. 10.2. S and S′ are moving along a common X = X′ axis with a relative uniform speed v. S′ is moving to the right of S. The inertial observers have metre sticks which are compared and checked; their clocks also are calibrated and synchronised. Let the primed coordinates refer to the frame S′ and the unprimed ones to the frame S. When the observers meet, let x = x′ = 0 and t = t′ = 0. After a lapse of some time, the observers O and O′ measure the coordinates of a point P in space as (x, y, z) and t and (x′, y′, z′) and t′, respectively. A large set of such numbers determines the motion of a body P for each of the observers. Naturally it is of interest to know how the sets (x, y, z), (x′, y′, z′) and t, t′ are correlated with each other. From Fig. 10.2, we have x′ = x − vt; y′ = y and z′ = z. The above equations are called *Galilean transformations*. The equations tell us how to transform the space coordinates of a point in space from one inertial frame to another.

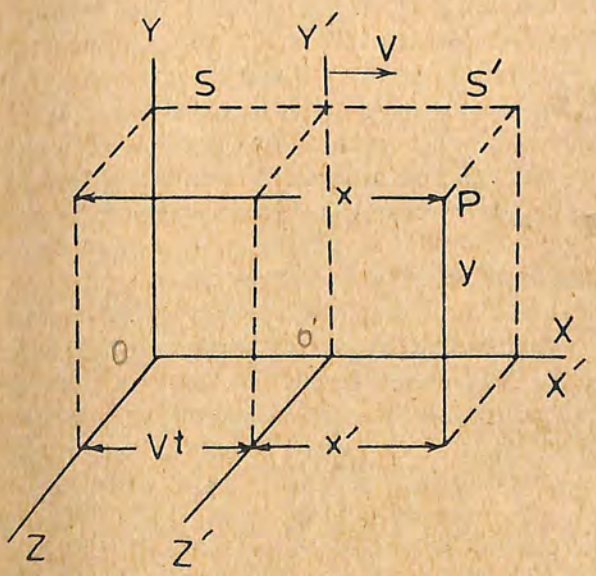


Fig. 10.2 Two coordinate systems S and S′ with the relative uniform speed V between them.

Prior to 1905, physicists believed that time is absolute, i.e., the time intervals are independent of the inertial frames. Such a belief did not bring about any inconsistencies between the experimental results and the predictions of Newton's mechanics. Thus, the equation t′ = t is added to the above set to complete the sets of transformations.

These sets of transformations contain information that space and time intervals are the same in all inertial frames. This statement defined space and time as absolute. Thus, in Newton's mechanics, both time and space are treated as absolute. Newton's mechanics regarded that space existed without any reference to objects. Space was absolute and at rest. All objects, like stars, etc., moved in this space. A frame of reference attached to space constituted an inertial frame at rest with reference to which all relative motion had to be determined.

10.4 Newtonian Relativity Principle

The complete set of Galilean trans-

formation equations is:

$$\left.\begin{aligned} x' &= x - vt \\ y' &= y; z' = z; t' = t. \end{aligned}\right\} \quad \ldots (10.1)$$

The body P in space (Fig. 10.2) moves in time. The sets (u_x, u_y, u_z) and (u'_x, u'_y, u'_z) represent the velocity components of the body for the observers O and O′, respectively. The relation between the sets is obtained by differentiating the equations 10.1. The relations are:

$$\left.\begin{aligned} u'_x &= u_x - v \\ u'_y &= u_y; u'_z = u_z \\ \text{or } \mathbf{u'} &= \mathbf{u} - \mathbf{v}. \end{aligned}\right\} \quad \ldots (10.2)$$

This is the velocity addition theorem in mechanics which specifies how velocities are to be added. In a similar manner, the relations between the acceleration components of a body P in the inertial frames S and S′ are:

$$\left.\begin{aligned} a'_x &= a_x; a'_y = a_y; a'_z = a_z \\ \text{or } \mathbf{a'} &= \mathbf{a} \end{aligned}\right\} \quad \ldots (10.3)$$

The above equations tell us that the acceleration but not the velocity of a body is the same in all inertial frames and acceleration is an absolute quantity. Absolute quantities are also referred to as invariants; here, they are invariants under Galilean transformations.

Mass is an important concept in Newton's mechanics and is a property of all material bodies. From experience, it appears reasonable to assume that mass is an absolute quantity and does not depend on the inertial frames. In other words, mass is invariant under Galilean transformations. Newton's second law defines force. The invariance of mass and acceleration makes force also to be an invariant. Thus, we have established that Newton's laws on which the framework of mechanics was built are the same in all inertial frames which is just a restatement of the Principle of Relative Motion.

Similarly, the principles of conservation of energy and momentum must be valid in all inertial frames. Thus, Newtonian relativity can be considered as a statement of invariants under Galilean transformations.

10.5 Nature of Light

The validity of mechanics is verified for the motion of objects whose speeds are very small compared to the speed of light. It is of importance to extend the validity of mechanics to a still higher speed as that of light. The outstanding feature of light is its tremendous speed (3×10^8) m/s. Macroscopic material bodies move with much smaller speeds; the earth's speed is 30 km/s around the sun. Light and sound exhibit certain similar characteristics. Both have wave characteristics. Sound requires a material medium for its propagation and cannot travel in a vacuum. Based on the above facts, scientists in the nineteenth century concluded that light must also require a material medium for its propagation; this material medium was named as *ether*. As light was seen from distant stars, it was natural for scientists to assume that ether pervaded all space in the universe. Ether as a medium must have certain observable properties. The propagation characteristics of light must depend on this medium.

According to the velocity addition theorem in mechanics (equation 10.2), the speed of light should not be the same in all inertial frames with the relative speed v, i.e., the speed of light is a variant under Galilean transformations. Scientific thinking at the end of the nineteenth century was that if the earth moved relative to ether, then the speed of light should be direction dependent, just like the directional dependence of a boat's speed in a river. The motion of the earth relative to

ether is so small as compared with c, the speed of light, that very small dependence of c on the direction was expected.

Michelson and Morley (1887) were the first to attempt the detection of any change in the speed of light due to the motion of the earth relative to ether.

10.6 The Michelson and Morley Experiment

The aim of the investigation was to detect the earth's relative motion in ether by measuring the speed of light in different directions on the earth's surface. The results were so startling and unexpected that various groups repeated this experiment with improved accuracy for several decades. A schematic diagram of Michelson interferrometer is shown in Fig. 10.3. A monochromatic light beam of a wave length λ from a source S is incident on a semi-silvered optical glass slab A. The glass slab makes a 45° angle with respect to the beam's axis. Part of the light (beam-1) is reflected from A and travels to the mirror D at a distance L_2 from A, and part of the light (beam-2) travels to the mirror B. The distance AB is L_1. The separated beams get reflected from the mirrors B and D and enter the telescope T. The speed v of the earth's relative motion to ether is also shown in the diagram. A given point in the field of the telescope will be bright or dark, depending on whether the path difference of the beams (ADAT and ABAT) is integral multiple or half integral multiple of the wave length. Thus, a pattern of alternately dark and bright fringes will be seen in the field of the telescope.

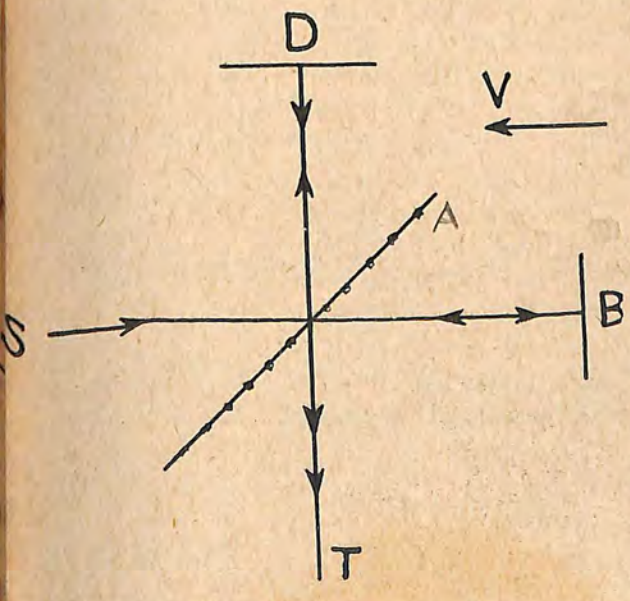


Fig. 10.3 The Michelson interferrometer.

The time difference for the beams 1 and 2 to travel up and down and across the ether wind is Δt. If the whole equipment is rotated through 90° about its axis, then the beam 1 has to travel perpendicular to the direction of the motion of the earth and the beam 2 along and opposite to the direction of the motion of the earth. The direction of the ether wind remains the same as before. The time difference for the beams 1 and 2 to travel now is $\Delta t'$. Therefore, the rotation of the interferometer changes the difference in the transit times of the beams 1 and 2 by $\Delta t' - \Delta t$. Since the transit time is different before and after the rotation of the equipment through 90°, there should be a fringe shift across the wire of the telescope. Exact calculations indicated that 4/10th of a fringe shift should be observed in the Michelson and Morley experiment. Michelson's interferometer is so sensitive that 1/100th of a fringe shift could be detected. Observations were made during days and nights (as the earth spins about its axis), and during all seasons of the year. The expected fringe shift was not observed. The final conclusion was that there was no fringe shift at all. This unexpected result is inconsistent with the Newtoniam relativity principle.

10.7 Special Theory of Relativity*

We have seen from the Michelson and Morley experiment that ether has no influence on c and it is the same in all directions in an inertial frame. According to the Newtonian relativity, if absolute space was identified with ether at rest, the speed of light c should be direction dependent on the motion of the earth. Since no such direction dependence for c was detected, the concepts of ether and absolute space became meaningless. Thus, in Einstein's view, there is no absolute space and all motion is relative.

Having seen that space is relative, Einstein questioned whether time also is relative. To realize fully the implications of the statement that time is relative, let us consider the following experiment. Let us consider a moving laboratory, e.g. a railroad car ABCD moving with the velocity v. (Fig. 10.4). An observer 1 at O_1 on one side of the car looks at a match struck on the other side (say, at O). To the observer 1, the light from the match seems to have come straight across the car (i.e., along the path OO_1). But an observer 2 at O_2 who is standing on the platform sees the light from the match takes a longer path before it reaches O_1. The observer 2 sees the observer 1 moving along the track from O_1 to O_1', and the light crossing diagonally from the point where the match was struck to the point O_1'. The two paths which the observers 1 and 2 see are, respectively OO_1 and OO_1'. Since the speed of light is to be the same for both the observers, it follows that the time interval between the two events (i.e., striking the match at O and the arrival of the signal at the observer 1) is longer for the observer 2 (the bystander) than that for 1 (the traveller). Time passes faster for the bystander than for the traveller. Thus the rate, at which time passes for the two inertial observers, is not the same, that is, time is relative.

These new concepts of space and time have prefound effect on the dynamics of

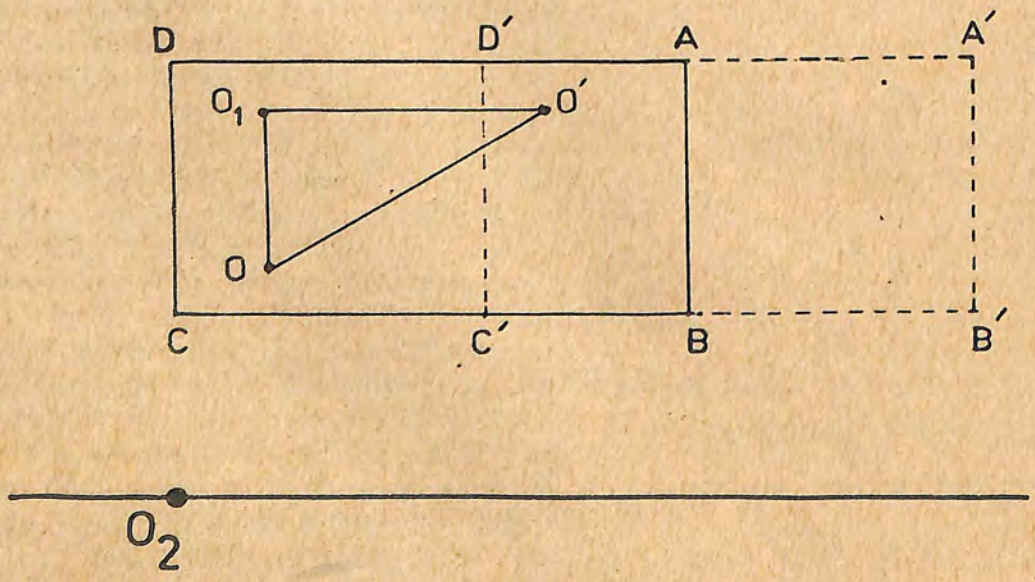


Fig. 10.4 Time runs differently for two inertial observers—1 at O_1 and 2 at O_2. Observer 1 is sitting in the rail car ABCD moving with the velocity v and observer 2 is standing at O_2 on the platform.

* In general, the relative motion between observers may be accelerated or decelerated and need not be inertial. If we confine to the inertial systems, then the special theory of relativity has to be used. For the non-inertial systems, Einstein in 1916 had proposed a theory now known as General Theory of Relativity.

bodies, i.e., mass is no more an absolute quantity as observed in Newton's mechanics. Einstein turned his attention to the concept of mass and the role it should play in the theory of special relativity.

It can be shown that the momentum P attained by a body of mass m under the action of the finite constant force F for the time t, is P = mu = Ft. Imagine a uniform electric or gravitational field pervading all space. An electron or a body in these fields would automatically be speeded up without limit in course of time. Since F is finite and constant u is proportional to t, only when m is a constant, i.e., m does not depend on its speed. This is one of the implicit assumptions of the Newtonian relativity, and, as a consequence, as t grows bigger and bigger ($t \to \infty$), u also attains values larger and larger without limit. But the speed of a physical object can be at most the speed of light. Thus, an upper limit on P is mc, a finite and definite value. Where is the flaw in the argument? Mass can be regarded as a constant as long as u is proportional to t. The upper limit of u is c, this value for a physical object is attained only when t is infinite in the above equation. In other words, u is not proportional to t for large values of u. Thus, when t is infinity u cannot be infinite, the only quantity that can assume an infinite value in the above equation is m. Thus, we are led logically to the conclusion that the mass of an object is not a constant but is a function of its own speed. What does this mean? Mass is finite and has a constant value m_o, when u = 0, but is infinity when $u \to c$ therefore, it is clear that mass varies as a function of its velocity. From detailed considerations, Einstein derived an expression, where m_o is the rest mass (u=0) of the object and m is the mass of the object when its speed is u.

$$m = m_o/\sqrt{1 - u^2/c^2} \quad \ldots (10.4)$$

We are familiar with the fact that one form of energy can be converted into another form. The kinetic energy of bodies (macroscopic as well as microscopic) is associated with their speed. As mass also is shown to be a function of its speed, would it not be natural to regard mass also as a form of energy? Einstein showed that mass itself is a form of energy and there is equivalence between mass and energy. The energy stored in an object of mass m is

$$E = mc^2 \quad \ldots (10.5)$$

where c is the speed of light and E is the energy. $E = mc^2$ established equivalence between mass and energy. This is known as the principle of equivalence between energy and mass. This principle has important applications in nuclear technology, popularly referred to as 'atomic energy'. What is nuclear energy? It is nothing but the transformation of mass energy stored in nuclei, like uranium into other forms. The transformation occurs in an assembly called a *nuclear reactor*.

This Unit is a re-examination of the concepts of space, time and mass based on the observations of physical bodies with the speed up to but less than the speed of light. Unlike Newton who regarded space and time as independent entities and absolute, Einstein viewed space and time as dependent entities and relative. The later concepts of space and time, which are diametrically opposite to those defined in the Newtonian relativity, are basic foundations to restructure the principles of relativity which are valid in all branches of physics.

EXERCISES

10.1 If it is desired to measure the length of an object, how would you estimate it when (i) the object is stationary, and (ii) it is moving with the uniform speed, relative to your frame of reference.

10.2 A person at the North pole defines his up-and-down directions. Similarly, another person at the South pole does the same. The up-and-down directions are not unique for both. Why?

10.3 A stone is thrown vertically up by a person and it returns to its starting point. Does this experiment prove that the earth revolves round the sun at 30 km/sec.?

10.4 Tabulate the following as variants and invariants under the Galilean transformations: Space, Time, Mass, Velocity, Acceleration, Force, Momentum, Kinetic Energy, Impulse, Work, Power, Torque and Angular Momentum.

10.5 In a reference frame S, two marbles, each of mass 3 kg, move along the x-axis with the speed 4 m/s (u_1) and -3 m/s (u_2), respectively. Use the principles of conservation of energy and momentum and compute the speed of the marbles after collision. S′ is moving with the speed 3 m/s, as in Fig. 10.2. Show that momentum and energy are conserved in the S′ frame also.

($u'_1 = -3$ m/s, $u'_2 = 4$ m/s)

10.6 Why should Michelson and Morley repeat their experiment during nights and days and during all seasons of the year?

10.7 If one gram of a substance is fully converted into heat energy, how many calories of heat would be produced? (Hint: $E = mc^2$)

(2.1×10^{13})

10.8 Calculate the mass of a body when its speed is 0.8 times the speed of light. Assume the rest mass m_0 of the body to be one gram. What will be its momentum at this speed ?

10.9 If 1000 kg of water is heated from 0° C to 100° C, calculate the increase in the mass of the water.

UNIT 11

Thermodynamics

We are already familiar with many topics in heat such as properties of gases, change of states, calorimetry, etc. We have yet to study many other topics such as heat in relation to mechanical work, radiation by hot bodies, pyrometry, etc. All such topics fall under the scope of thermodynamics. The detailed study of these topics is beyond the scope of this Unit. However, we will now try to study some general principles of thermodynamics and their applications to some familiar situations. So we begin with the development of topics which are pre-requisites for this study.

11 1 Temperature

Consider two bodies, A and B. Let body A be such that it feels cold to the hand, and body B be such that it feels hot to the hand. We then say that the body B is at a higher temperature than the body A. Let the two be kept in contact with each other. After some time we will find that both A and B give the same temperature sensation, and we then say that the two are in thermal equilibrium with each other.

The test for the thermal equilibrium of two bodies is to use a third body which should be such that it does not appreciably disturb the thermal state of either of the bodies.* Suppose body A is in thermal equilibrium with a body C, say a thermometer, and body B is also in thermal equilibrium with body C, the thermometer; then A and B are in thermal equilibrium with one another. This statement is known as the Zeroth Law of Thermodynamics.

11.2 Heat Energy

Drop a hot body in a vessel containing cold water. The water warms up till the temperature of the two becomes the same. This change in temperature is due to the fact that energy flows from the hot body to the cold body. This flow of energy from one body to the other, which takes place by virtue of their temperature difference alone, is called heat energy, or just heat for short. The hot body and the cold water, if considered separately, do not have any heat energy though they do have energy. Heat energy becomes evident only when energy flows from one to the other. When the two attain a common temperature; the quantity of heat energy transferred will be zero and the two systems will be in thermal equilibrium. In

*This means the heat capacity, i.e., mass into specific heat, of the test body should be much smaller than that of the bodies whose temperature it has to measure.

thermodynamics, an adiabatic* process is one in which there is no transfer of heat energy.

Conventionally, the quantity of heat energy transferred to a body is said to be positive (+Q), and that transferred from a body is said to be negative (−Q).

Till recently, the unit of heat energy was the calorie. It is usually defined as the quantity of energy required to raise the temperature of 1 g of water through 1°C. However, it has recently been internationally agreed to express all energy changes in the energy unit, the joule (J). A joule is defined as the work done when a force of one newton acts through one metre. The relation between the two is : 1 calorie =4.18 joules. If the amount of heat energy supplied to a system is 5 calories, then Q= 20.9J

11.3 Work

Consider a cylinder having a movable piston (Fig. 11.1). Let it contain one gram mole of a perfect gas. If the pressure exerted by the gas be P and the area of cross-section of the cylinder be A, then the force exerted by the gas on the piston will be PA. If due to this force the piston moves an infinitesimal distance d*l*, then the corresponding work done will be

$$dW = PA\ dl$$
$$= PdV$$

where dV is the infinitesimal volume change A d*l*. What will be the work done if the piston moves a finite distance ? Consider the case when the piston steadily moves through a finite distance and the volume of the gas increases from V_1 to V_f. If the process is carried out slowly, W will be given by

$$W = \int_{V_1}^{V_f} PdV \qquad \ldots (11.1)$$

Is this work done a positive or a negative quantity ? We make use of the following convention.

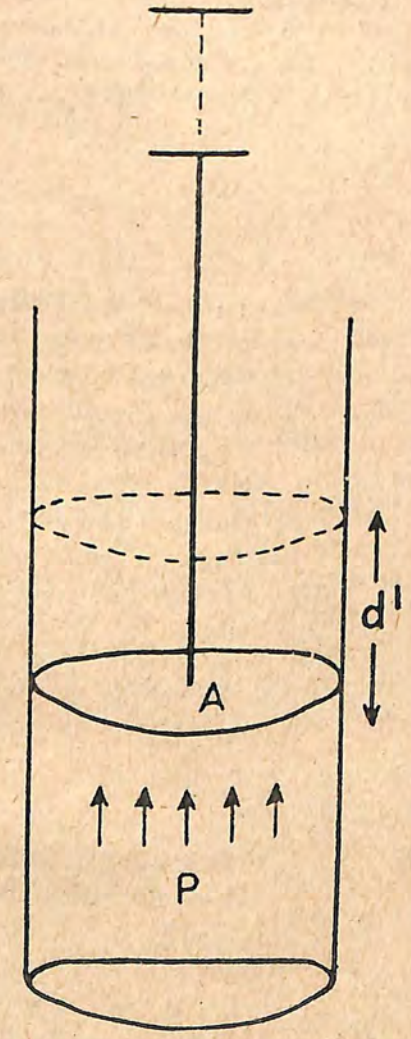


Fig. 11.1 Expansion of gas inside a cylinder

*a=not, dia=through, bates=heat
adiabatic is a Greek word meaning heat not passing through.

In thermodynamics, an object or a group of objects under consideration is referred to as a system and everything external to the system is called the surroundings.

So in our discussion the cylinder with the piston containing gas constitutes the system. Conventionally, the work done by a system is taken to be positive.

Let us consider the converse process. Suppose the gas is compressed by adding weights on the piston; then the work done W' will be

$$W' = \int_{V'_i}^{V'_f} P' \, dV$$

where P' is the pressure of the gas. In this case the work is done on the system and conventionally W' will be negative.

The above concepts are illustrated in the following example.

EXAMPLE 11.1

Three moles of an ideal gas kept at a constant temperature of 300K are compressed from a volume of 4 litres to a volume of 1 litre. Calculate the work done in the process.

$$W = \int_{V_i}^{V_f} P dV$$

Since PV=nRT, for an ideal gas, we get

$$W = \int_{V_i}^{V_f} \frac{nRT \, dV}{V}$$

$$= nRT \log_e \frac{V_f}{V_i}$$

Since n=3 mole, T=300 degree,

R=8.31 J/mole deg.

and $\log_e x = 2.3 \log_{10} x$ we get

$$W = 3 \times 8.31 \times 300 \times 2.3 \log_{10} 1/4$$

$$= -10320 \text{ J.}$$

The negative sign indicates that work is done on the system.

In the above we considered the process of expansion of the gas and calculated the work done in the process. We used the relation PV=nRT. The analytical method would fail where the relationship between P and V is not known. In such cases the same can be done graphically as described below.

Indicator diagram

Consider the system in which the gas exerts pressure on the piston of the cylinder (Fig. 11.1). Let the volume of the gas increase very slowly. In order to represent the process graphically, let the volume V be plotted along the abscissa and the pressure P along the ordinate. By plotting the values of P and V for successive positions of the piston, the process can be represented by a curve as in Fig. 11.2.

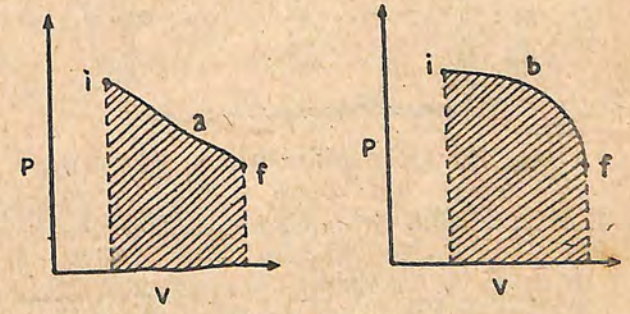


Fig. 11.2 Indicator diagram

The area under the curve gives the work done in the process, PdV. Such a diagram in which pressure is plotted along the ordinate

and volume along the abscissa is called a P-V diagram, or an indicator diagram.

11.4 Work and Heat

We have seen that both heat and work are measured in the same units, namely joule. So the question arises what is the relation between the two quantities. Joule was one of the early scientists who experimentally studied this. He proceeded as follows.

A rod to which some paddles are attached was mounted in a cylinder containing water. The rod could be revolved by a thread and pulley arrangement as shown in Fig. 11.3. The

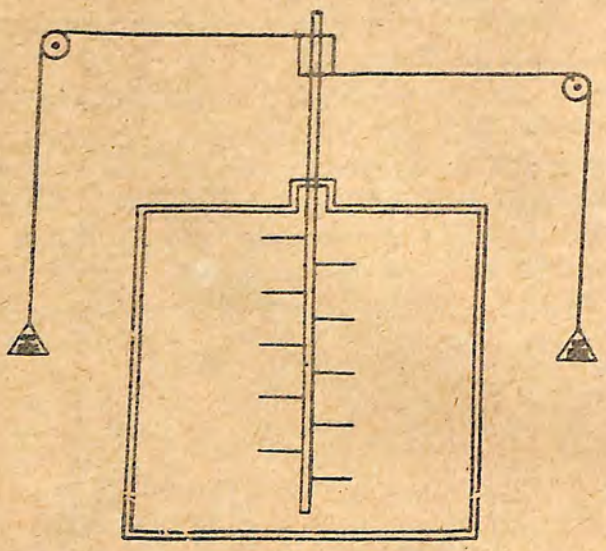


Fig. 11.3 Apparatus for Joule's experiment

water got heated on revolving the rod. The quantity of heat gained by the water was calculated by noting the rise in temperature.

The quantity of work done in revolving the rod was also calculated, by noting the distance through which the weights fell. By applying necessary corrections for the heat energy lost in radiation, frictional losses, etc., Joule concluded that a definite relation exists between the work done and the heat gained. Mathematically, he expressed this relation as

$$W = JH \qquad \ldots(11.2)$$

where W is the work done in joules, H is the heat gained in calories and J is a constant. J is called the mechanical equivalent of heat and from his experiment Joule determined its value as 4.18 joules per calorie.

So we see from above that heat and work are inter-related and all of the work done can be converted into heat. What about the converse process? Can all of heat supplied be converted to work? We will discuss this later.

11.5 Internal Energy

We know that work and heat are both some form of energy and the two are inter-convertible. Let us consider the case of compression of a gas enclosed in a cylinder with a moving piston. When the piston is pushed by doing some work, the gas gets compressed and its temperature rises. If the walls of the system are thermally insulated (Fig. 11.4) no heat will be transferred i.e. the process will be adiabatic. The work done in the system gets converted into another form of energy which is called the internal energy of the system and the total energy remains conserved. Most of the internal energy is due to the kinetic energy arising from the motion of the molecules inside the system. When a thermodynamic system changes from one state to another, its internal energy also changes. The changes in the internal energy when the system changes from one state to another depends only on the initial and final states of the system. Internal energy is denoted by the symbol U. We can summarise the above by stating that in an adiabatic process the mechanical work (—W) done on the system

increases the internal energy of the system from U_i to U_f such that

$$-W = U_f - U_i \quad \ldots(11.3)$$

Internal energy of a system changes during certain thermal processes i.e. during change of phase, say from liquid to vapour. We will discuss it while studying first law of thermodynamics.

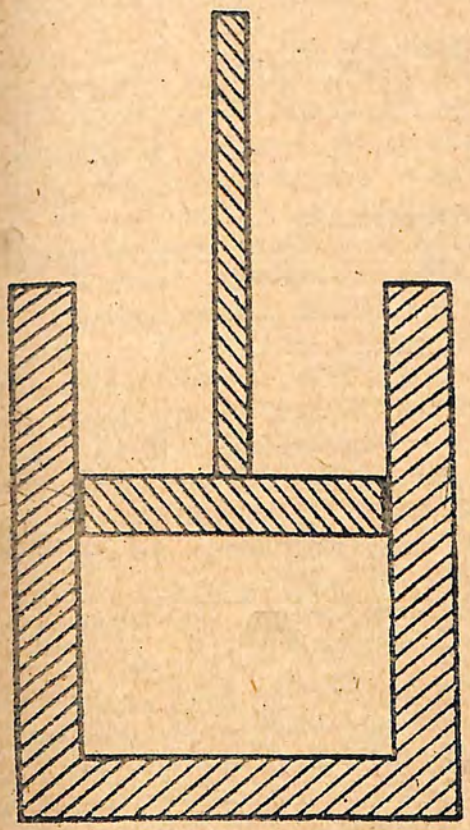


Fig. 11.4 Apparatus for adiabatic experiment

In the above some quantities commonly used in thermodynamics have been briefly introduced. Besides, we learnt that work can be converted into heat and that a definite relation exists between the two. We will now proceed to discuss the laws of conversion of heat into work.

11.6 First Law of Thermodynamics

Consider a gas inside a cylinder with a movable piston. Let all the walls of the cylinder except the bottom be thermally insulated. Let the bottom be brought into contact with a hot body or a burner (Fig 11.5) so that it absorbs a quantity of heat Q.

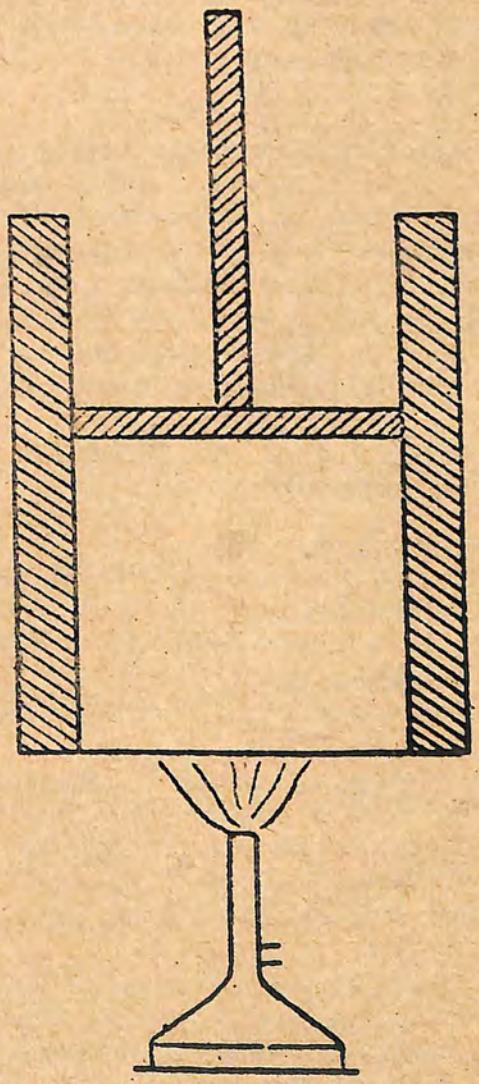


Fig. 11.5 Absorption of heat by a gas enclosed in a cylinder

On absorption of heat the temperature of the gas will rise and it will expand. In the process the piston will be pushed upwards and in doing so the gas does some work (W), and the internal energy of the gas increases from U_1 to U_f.

The amount of heat energy absorbed by the gas (+Q) is used by it in doing external work (+W) and increasing its internal energy by $(U_f - U_1)$. Since the total energy has to be conserved we can write

$$Q = W + (U_f - U_1) \quad \ldots(11.4)$$

The above equtaion mathematically represents what is known as the first law of thermodynamics. It can be stated in words as follows. If heat is supplied to a system which is capable of doing work then the quantity of heat absorbed by the system will be equal to the sum of the external work done by the system and the increase in its internal energy.

11.7 Applications of the First Law of Thermodynamics

(a) Boiling process

Let us apply the first law of thermodynamics to the boiling process. We know that on adding heat a substance changes its phase from liquid to vapour. For example, on being heated, water boils at 373K at atmospheric pressure.

Consider the vaporization of a mass m of a liquid at constant temperature and pressure, P. Let V_1 be the volume of the liquid and V_f that of the vapour. The work done (W) by the liquid in expanding will be given by $W = P(V_f - V_1)$.

Let the heat of vapourization be L. It represents the heat needed per unit mass to change from liquid to vapour phase at constant pressure and temperature. The heat absorbed Q by the liquid will be given by $Q = mL$

By applying first law of thermodynamics to the process, we get

$$Q = U_f - U_1 + W$$

or $$mL = U_f - U_1 + P(V_f - V_1)$$

Knowing m, L, P, V_f and V_1 gain in internal energy can be easily calculated.

(b) Specific heat relation

While studying the Unit on Kinetic Theory of Gases we discussed the equation $C_p - C_v = R$. We will now see how this equation can be obtained by thermodynamic considerations.

The specific heat of a substance is defined as the heat per unit mass required to raise its temperature by one degree. Accordingly, the specific heat of one mole of a gas is called molar specific heat C.

Consider n moles of an ideal gas. If its temperature is raised by an amount dT at constant volume then the heat transferred is $nC_v dT$ where C_v is the molar specific heat at constant volume. Since there is no change in the volume the work done in the process will be zero.

By applying first law of thermodynamics

$$dQ = dU + PdV$$

We get $$nC_v dT = dU$$

Next, instead of keeping the volume constant let the same gas be heated by keeping the pressure constant. The heat required to change its temperature by the same amount will now be $nC_p dT$, where C_p is the molar specific heat at constant pressure. The work done in expansion will be PdV. By applying the first law to the process we get

$$nC_p dT = dU + PdV.$$

We can assume that the change in the internal energy of an ideal gas is entirely due to change in its kinetic energy. We have learnt in the Unit on Kinetic Theory of Gases that the change in the kinetic energy of an ideal gas depends only on the change in the temperature.

In the above, since the change in temperature dT is the same, dU will also be the same in both the cases. Substracting one from the other, we get

$$nC_pdT - nC_vdT = PdV$$

Since for an ideal gas $PV=nRT$, we can write $PdV=nRdT$. So we have

$$n(C_p - C_v)\,dT = nRdT$$

or $$C_p - C_v = R$$

11.8 Conversion of Heat into Work—Heat Engine

Consider the familiar case in which steam is used to do work. The working substance is steam and the heat engine i.e. the mechanical device with the help of which steam undergoes a cycle, is the steam-engine. The heat is taken from the boiler and part of it rejected to the surrounding atmosphere.

In general a system which is used to convert heat into work, absorbs a quantity of heat Q_1, performs an amount of work W and returns to the initial state after rejecting some heat Q_2 (Fig. 11.6). The reservoir of heat from which it absorbs heat is at a higher temperature and is called the source. The reservoir of heat to which it rejects heat is at a lower temperature and is called the sink. The magnitude of the source and the sink is so large that their temperature remain unchanged because of any heat supplied to them or removed from them.

While discussing about the work done by an engine it is useful to consider the work done in each cycle. In thermodynamics by a cycle we mean an operation satisfying the following conditions. The system starting from an initial state is brought back to the same state after undergoing through a series of processes. The system does not retrace its path and its indicator diagram is a closed curve and not a line. All the properties of the system remain unchanged during the process.

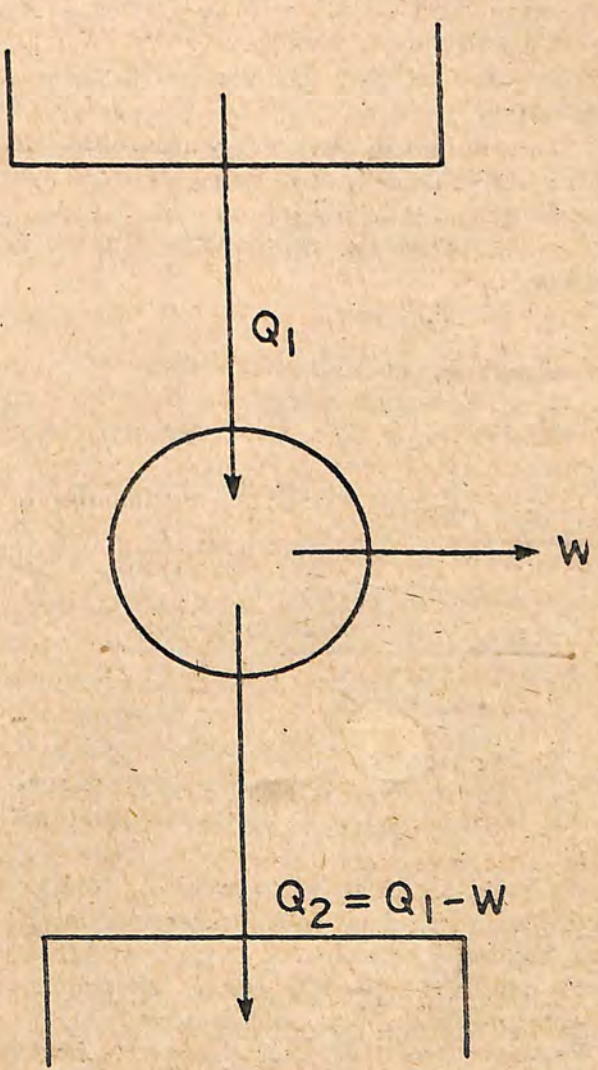


Fig. 11.6 Heat engine

The work done by a heat engine will be larger if more and more work is done in each cycle of the engine. More precisely we say that the work done by an engine is larger if its thermal efficiency is higher. Thermal efficiency

of an engine, is defined as

$$\text{efficiency} \quad (\eta) = \frac{\text{work output } (W)}{\text{heat input } (Q_1)}$$

$$\eta = \frac{W}{Q_1} \qquad \ldots(11.5)$$

We can substitute for W using the first law of thermodynamics. Since the system returns to its initial state after completing the cycle there is no change in its internal energy. So we can equate the change in the heat energy of the system to the amount of work done by the system.

or $Q_1 - Q_2 = W$

So we can write $\eta = \frac{Q_1 - Q_2}{Q_1}$

or $\eta = 1 - \frac{Q_2}{Q_1}$ $\qquad \ldots(11.6)$

The value of η is found to be always less than unity. Why is it so ? This we will discuss below in our study of second law of thermodynamics.

11.9 Second Law of Thermodynamics

We have seen that heat flows from a body at higher temperature to a body at lower temperature. But it is common knowledge that heat does not flow from a body at a lower temperature to the one at a higher temperature. Why ?

We have seen that when work is done in rotating a paddle kept in a beaker containing water, the water gets heated. But why is it that when a paddle is kept in a beaker containing hot water no mechanical work is done ?

We know that heat engine converts heat into work. But why is it that the efficiency of an engine is always less than unity ?

All these questions lead us to second law of thermodynamics.

We know that if an engine is to work continuously and more and more work has to be done, either the working substance should be inexhaustable or, it should work in a cycle. Since the first possibility is not practicable we will have to consider an engine working in a cycle. In practice it is seen that no engine ever developed, working in a cycle, extracts heat from a hotter body and converts all of it into work, without rejecting some heat to a sink at lower temperature. All of the heat extracted from the source has never been converted into work. These observations were generalised by Kelvin and also by Planck. Their statements can be combined into one equivalent statement, as "It is impossible to construct an engine, operating in a cycle, which will produce no effect other than extracting heat from a reservoir and performing an equivalant amount of work." This statement is known as Kelvin Planck statement of the second law of thermo dynamics.

The above law implies that the working substance working in a cycle cannot convert all of the heat extracted into work. It has to reject some amount of heat to the sink. So in order to convert heat into work it is essential to have both a source and a sink. Since all of heat extracted is never converted into work it follows that the efficiency of an engine is never unity. Of course η can never be more than one because this implies that Q_2 is negative i.e. the working substance extracts heat both from the source and the sink and rejects no heat, which we know by second law, is absurd.

11.10 Second Law and Refrigerator

We saw that in a heat engine the working substance, working in a cycle, absorbs some heat energy from a hot reservoir, converts part of it into work and transfers the rest to the sink. In practice, it is possible to construct a device which acts in the reverse way. In this the working substance absorbs some heat from a sink at low temperature and on some work being done on it a larger amount of heat

is rejected to the reservoir at higher temperature. Such a device is called a refrigerator, and its working substance is called the refrigerant. The working of a refrigerator can be briefly described as follows.

Let the quantity of heat absorbed by the refrigerant at a lower temperature be Q_2 (Fig. 11.7). The work done on the refrigerant be W.

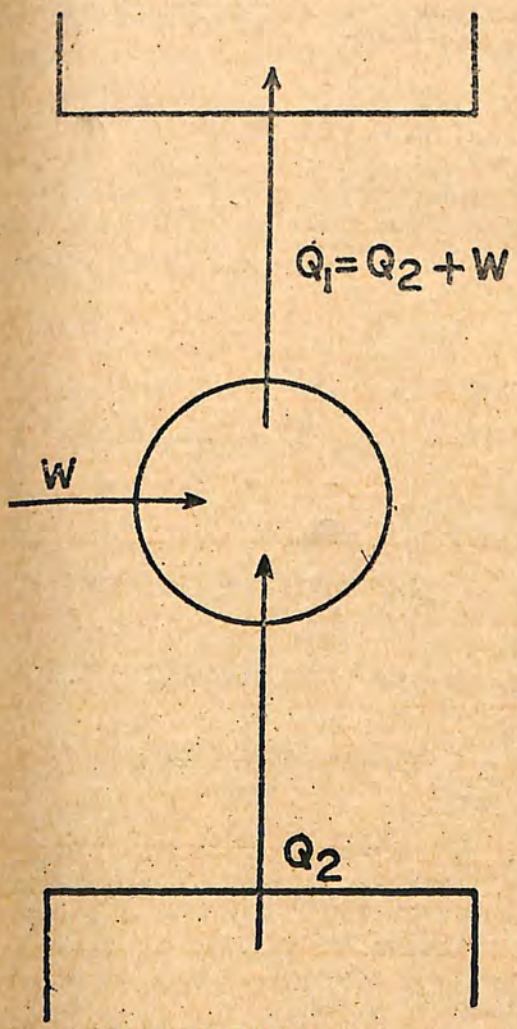


Fig. 11.7 Refrigerator

The heat rejected by the refrigerant be Q_1. The substance is working in a cycle such that there is no change in its internal energy. Substituting these quantities in the mathematical form of the first law equation (Eq. 11.4), we get

$$Q_2 - Q_1 = -W$$

or $$Q_1 = Q_2 + W \qquad \ldots(11.7)$$

This equation implies that it is always necessary to do work on the substance in order to transfer heat from a sink to the source.

In household refrigerators the work is done by an electric motor. The working substance commonly used in it is freon. It absorbs heat from materials kept inside the refrigerator and rejects it to the surrounding air which is at a higher temperature.

11.11 Reversible Process

In thermodynamics a process is said to be reversible if it can be retraced in the opposite direction such that the system and the surroundings pass through exactly the same states at each stage as in the direct process. After the conclusion of a process, the system and the surroundings are restored to their initial state without producing any change in either of them. A reversible process has to be done very slowly such that it satisfies the following requirements at every stage of the process.

(a) The system is in mechanical equilibrium, i.e. there is no unbalanced force in its interior or between the system and the surroundings.

(b) The system is in thermal equilibrium, i.e. all parts of the system and the surroundings are at the same temperature.

(c) The system is in chemical equilibrium, i.e. it does not spontaneously change the internal structure due to chemical reaction, diffusion, etc.

Since any system which satisfies the above three conditions is said to be in thermodyna-

mic equilibrium the above conditions may be summarised as follows. A reversible process at every stage during the process should be in thermodynamic equilibrium. Besides of course a reversible process should be devoid of any dissipative effects such as frictional losses, etc.

Let us consider some common processes. Motion of a body in the ground is irreversible because the energy spent in overcoming friction is not recoverable. Similarly stirring of a liquid and conduction of heat are irreversible as some energy is lost due to radiation. We can generalise that most natural processes are irreversible. All chemical reactions are irreversible because they involve change in the internal structure of the constituents. The constituents cannot be restored unless there is energy transfer from outside.

11.12 Carnot Engine

We have learnt before that the efficiency of a heat engine is always less than one. This leads us to the question as to how to operate an engine so that its efficiency is maximum. Let us try some possible solutions. Suppose we have a heat engine which works isothermally and that the working substance is an ideal gas. Then on absorbing heat from the source, the gas will expand to do some work. In order to do more work, the size of the engine will have to be continuously increased and this is not practicable. Alternatively the engine must work in a cycle. But when the working substance is brought back to the initial state by doing external work the indicator diagram for the complete cycle will just be a line. Area enclosed in the curve will be zero and so no net work will be done in the process.

Next, let us suppose that the heat engine works adiabatically. On compression the temperature of the gas will rise. The compressed gas will do some work. If the gas is brought back to its original state along an adiabatic then $dQ=0$, $dU=0$ and from the first law of thermodynamics it follows that no net work will be done in the process.

So the question remains as to how to operate the engine in an efficient way. Sadi Carnot, a French engineer, worked on the idea that if the indicator diagram of an engine is made up of suitable combination of isothermals and adiabatics then more and more work can be done. He described how to carry out a set of operations so as to achieve the maximum possible efficiency in a heat engine. These set of processes constitute the Carnot Cycle. Any engine which operates by performing these processes is called a Carnot engine. Its working is briefly described below (Fig. 11.8).

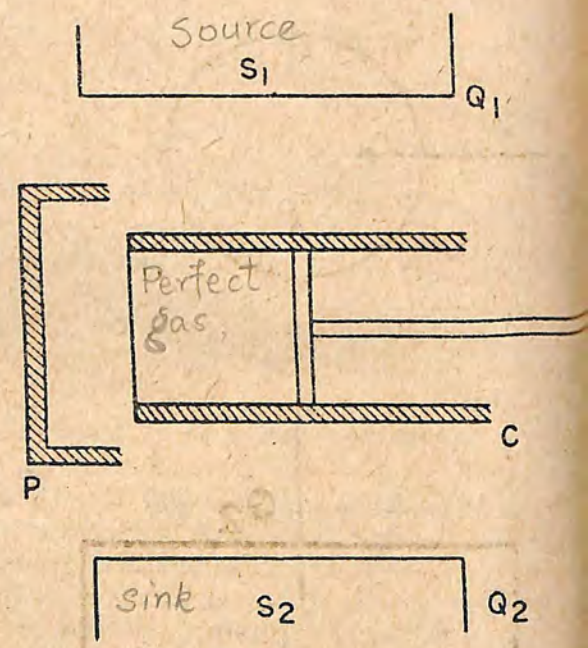


Fig. 11.8 Carnot engine

Let S_1 be the source at temperature θ_1, the sink at temperature θ_2, C be the cylinder of the engine fitted with a non-conducti

piston and P be a non-conducting cap which can be screwed to the bottom of the cylinder. Let all the sides of the cylinder, excepting its bottom be non-conducting.

Let the working substance inside the cylinder be a perfect gas. The behaviour of the working substance during the four successive stages of the operation is described in Fig. 11.9.

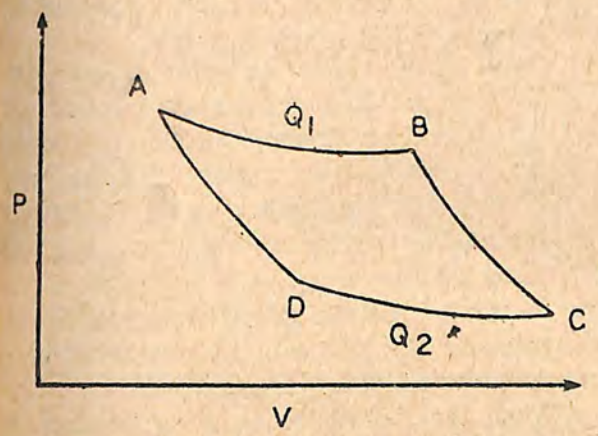


Fig. 11.9 Carnot cycle

In the first stage, the insulating cap is removed from the bottom of the cylinder. The gas is kept in thermal contact with the source at temperature θ_1. The initial pressure and volume of the gas are represented by the point A in the indicator diagram. Let the gas be allowed to expand such that the piston moves very slowly, the gas expands steadily and the process is reversible. As the gas expands its temperature tends to fall. But since it is kept in thermal contact with S_1, it will absorb a quantity of heat Q_1. So all throughout the process of expansion the temperature of the gas will remain θ_1. Incidentally any process in which the temperature remains the same throughout is called an isothermal* process. During this isothermal expansion let the representative point move from A to B.

At this stage when the representative point is at B the second stage of the operation begins. The insulating cap P be fixed on the bottom of the cylinder such that there is no heat transfer from the gas. But the piston will continue to move through some distance because of inertia. This motion is reversible. All through the process of adiabatic expansion due to inertial motion, there is no transfer of heat, to or from the gas. During an adiabatic expansion, temperature will fall and let the temperature of the gas at the end of the expansion be θ_2, the temperature of the sink. Let the representative point move during this adiabatic process from B to C.

At the point C the pressure of the gas is quite diminished. In order to enable the gas to do more work, it is necessary to bring it back to the original state represented by the point A. This compression is done very slowly during the third and the fourth stages.

During the third stage of the process, the cap P is removed from the cylinder and the gas is kept in thermal contact with the sink at temperature θ_2. The gas is compressed isothermally. Though during compression the temperature of the gas will tend to rise, it will remain constant at θ_2 because heat developed is transferred to the sink. Let the quantity of heat transferred to the sink be Q_2. During this isothermal compression the representative point will move from C to D.

At this point the fourth stage of the operation begins. The cap P is replaced on the bottom of the cylinder. The gas is then further compressed in a reversible process. Since the

* iso equal, thermo heat, isotherm is the locus of points representing states of a system which are in thermal equilibrium with one state of some other system or isotherm is a locus of points having equal temperature; isothermal-pertaining to isotherms.

process is adiabatic, temperature of the gas will rise. The gas is compressed till the original state is reached, i.e. till the representative point reaches the initial point A. Thus the cycle is completed. The work done by the engine during the cycle is given by the area enclosed by the curve. The gas can be taken through the cycle again and again and more and more work can be done.

11.13 Efficiency of a Carnot Engine

The work done by the gas during the cycle will be equal to the area under the closed curve ABCDA. It follows from the first law of thermodynamics that since the system is brought back to the initial state, there is no change in its internal energy. So this work done, W will be equal to Q_1-Q_2, i.e. the difference between the heat supplied to the system and the heat rejected by the system. So we can write

$$W=Q_1-Q_2$$

$\therefore$ the efficiency of the Carnot engine will be

$$\eta=\frac{\text{work done}}{\text{heat supplied}}=\frac{W}{Q_1}$$

$$=\frac{Q_1-Q_2}{Q_1}$$

$$\eta=1-\frac{Q_2}{Q_1}$$

By calculating the values of Q_1 and Q_2, it can be shown that

$$\frac{Q_2}{Q_1}=\frac{\theta_2}{\theta_1} \qquad \ldots(11.8)$$

Where θ_1 and θ_2 are temperatures on the perfect gas scale.

So we can write

$$\eta=1-\frac{\theta_2}{\theta_1} \qquad \ldots(11.9)$$

11.14 Reversibility of Carnot Engine

While describing the Carnot engine we considered four reversible operations AB, BC, CD and DA. Let us now discuss the reverse process, AD, DC, CB and BA. To start with let the working substance represented by the point A be allowed to expand along the adiabatic AD. During this reversible process its temperature will fall from θ_1 to θ_2.

When the representative point is at D the cap is removed and the gas is allowed to expand along the isothermal DC in a reversible process by keeping it in contact with the sink at θ_2. The gas absorbs a quantity of heat Q_2 and maintains its temperature θ_2.

When the representative point reaches C the cap is inserted and the gas is compressed along the adiabatic CB. During this reversible process the temperature of the gas will rise to θ_1.

When the representative point reaches B the cap is removed, the gas is kept in thermal contact with source and is further compressed along BA. In this reversible process, the gas rejects a quantity of heat Q_1 to the source.

In the above set of processes the engine takes heat (Q_2) from the sink and on some work (W) being done on it, it rejects heat (Q_1) to the source at higher temperature. The engine, therefore, acts like a refrigerator (Art. 11.10). The system and the surrounding have returned to the original state A. So it follows that Carnot engine is a reversible engine.

It can also be shown that no engine is more efficient than a Carnot engine and that efficiency of all reversible engines is the same as that of Carnot engine. In the above we have considered perfect gas as the working substance. However, it can be shown that efficiency of a reversible engine is independent of the nature of the working substance.

All engines used in practice such as steam

engine, diesel engine, etc., are irreversible and their efficiency is less than that of a Carnot engine.

11.15 Radiation

We know that there are three principal modes of transfer of heat energy, namely, conduction, convection and radiation. Conduction of heat through a material is caused by the transfer of heat from particle to particle and does not involve any motion of the molecules of the material itself. In convection heat is propagated through the material by the actual motion of the molecules of the material. In radiation heat is transferred from one body to the other even in the absence of any medium in the intervening space.

Radiation is only a means of transfer of energy by transverse electromagnetic waves. It is similar in nature to light waves, radio waves, etc. But in this unit we are mainly concerned with thermal radiation i.e. transfer of energy by electromagnetic waves in the infra-red region.

11.16 Emission and Absorption of Radiation

Radiation emitted by a body depends on the nature of its surface. A dull blackened surface is a better emitter of radiation than a silvered polished surface. Radiation emitted is directly proportional to the size of the surface. A body at higher temperature will emit more radiation than that at a lower temperature. The energy emitted by a body can be expressed in terms of the emissive power of the surface.

For a given temperature of a body if e_λ is the emissive power of the surface then the radiant energy emitted per square metre per second between wavelengths λ and $\lambda+d\lambda$ will be $e_\lambda d\lambda$.

By experience we know that a bright polished surface reflects most of the radiation falling on it and a rough black surface absorbs most of the radiation falling on it. Different surfaces absorb radiant energy differently. We define the absorptive power a_λ of a surface for a given temperature and wavelength λ as the ratio of the radiation absorbed to the radiation incident on it between λ and $\lambda+d\lambda$.

11.17 Black-body

If a body absorbs all the radiation falling on it without reflecting or transmitting any of it, then it is said to be black. In other words the absorptive power of a black-body is unity.

It is impossible to get a perfectly black body but some devices can be made to serve as perfectly black bodies. For example, a cavity in the form of a hollow sphere with its inside coated with black material and a small conical opening (Fig. 11.10), may be considered as a

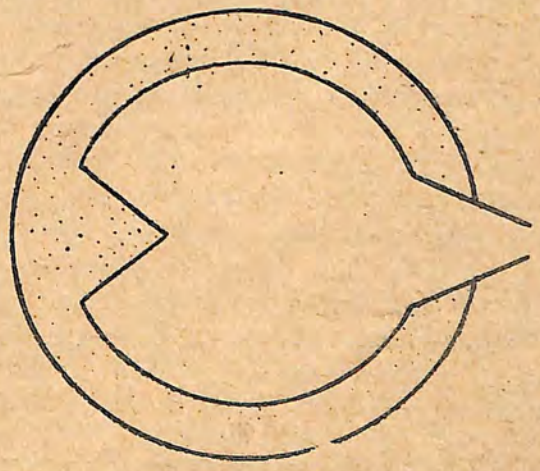


Fig. 11.10 A black-body

perfectly black body. It absorbs almost all the radiation falling inside it and we cannot

see what is inside the cavity. The intensity of the radiation coming out of the hole depends only on the temperature of the cavity.

The black body commonly used in practice consists of a metallic chamber made of brass or platinum and blackened inside. It is electrically heated using heating element. The radiation coming out of a small hole in the chamber is used for experimental purposes.

11.18 Kirchoff's Law

Consider a body which is in thermal equilibrium with the surroundings. It emits some radiation as well as absorbs some radiation falling on it. Since its temperature is supposed to be constant, its rate of emission must be equal to rate of absorption. For the given temperature of the body if e_λ is the emissive power of the surface then the energy emitted per square metre, per second, between λ and $\lambda + d\lambda$ will be $e_\lambda\, d\lambda$.

Let the energy falling on unit area per second within the same wavelength range and temperature be dQ. If for the given temperature the absorptive power is a_λ then the total energy absorbed per unit area per second will be $a_\lambda\, dQ$.

Since it is assumed that the body is in thermal equilibrium, we can write

$$e_\lambda\, d\lambda = a_\lambda\, dQ$$

$$\text{or} \quad \frac{e_\lambda}{a_\lambda} = \frac{dQ}{d\lambda}$$

We know that for any given temperature, dQ is a constant. So for a fixed range of wavelength $d\lambda$, dQ will be a constant. This implies that $\frac{e_\lambda}{a_\lambda}$ is a constant. This is true for any surface. If we consider a perfectly black surface then

$a_\lambda = 1$ and its emissive power will be E_λ. So we can write

$$\frac{e_\lambda}{a_\lambda} = \frac{E_\lambda}{1} = E_\lambda \qquad \ldots(11.10$$

We can generalise this equation as follows: "For any given temperature and wavelength the ratio of the emissive power to absorptive power is the same for all substances and is equal to the emissive power of a perfectly black body at that temperature". This law was first deduced by Kirchoff and is therefore known by his name. If the radiation covers a wide range of wavelengths then this law holds good for each wavelength considered, separately. This law implies that if a body is a good absorber of a particular wavelength of light then it should also be a good emitter of that wavelength. This law has many practical applications

11.19 Applications of Kirchoff's Law

If a polished metal piece with a black spot on it is heated to about 1200K and then suddenly transferred to a dark room the black spot will appear brighter than the polished surface. This is because the black spot being a better absorber is also a better emitter. Similarly, when a polished green glass plate which absorbs more of red region of the light when heated to a high temperature in a furnace and taken out glows with a red light.

Kirchoff's law is of great help to study the atmosphere of the sun. Its role is illustrated by the following experiment. When a white light is viewed through a sodium flame using a spectroscope in the laboratory two dark lines are observed in the continuous spectrum. Now if a flame tinged with NaCl is observed in the spectroscope two yellow lines are seen exactly in the same position where dark lines were observed before. This phenomenon is used to explain the dark lines in sun's spectrum.

The central body of the sun consists

glowing mass which emits a continuous spectrum without any dark lines. When this light passes through sun's atmosphere which is at a lower temperature, many elements present in it absorb their characteristic wavelengths and dark lines will be seen in those regions of the spectrum. These dark lines were first studied by Fraunhoffer and are therefore known by his name. They are characteristic of different elements present in sun's atmosphere. By studying emission spectra of different elements in the laboratory and comparing them with sun's spectrum, the elements in the sun's atmosphere can be identified.

11.20 Stefan's Law

The total energy emitted by a black body depends only on its temperature. The exact relation between the two was deduced by Stefan. He stated that the energy emitted per square metre of a black body per second is proportional to the fourth power of the absolute temperature, i.e.

$$E = \sigma T^4 \qquad \ldots(11.11)$$

where σ is a constant, called Stefan's constant and has the value 5.735×10^{-8} J/m^2/sec/deg^4. In order to calculate the net energy lost by a hot body the above law can be generalised as follows. "If a black body at absolute temperature T is surrounded by another black body at absolute temperature T_o, the amount of energy E_{net} lost per second per square metre by the body at higher temperature is given by

$$E_{net} = \sigma(T^4 - T_o^4) \qquad \ldots(11.12)$$

The theoretical proof of the above equation (11.11)) was later given by Boltzmann and so the law is also called as Stefan-Boltzmann law. This law is commonly used in estimating the temperature of hot bodies as described in the article on pyrometers.

11.21 Experimental Study of Black-body Radiation

The distribution of energy of the radiation emitted by a black body among the different wavelengths of the spectrum has been experimentally studied. The details of such measurements carried out by Lummer and Pringshein are briefly described here.

The source of black body radiation is an electrically heated chamber. The temperature of the source is measured using a thermo-couple. The radiation is dispersed into a spectrum by using a fluorite prism. The wavelength at different regions of the spectrum can be calculated by the known formula for dispersion. The intensity of spectrum at different regions is measured using a sensitive platinum resistance thermometer called the bolometer. The electrical resistance of the resistor in the bolometer changes with the temperature, and indicates the emissive power e_λ of the source between the wave length λ and $\lambda + d\lambda$.

The distribution of energy of the spectrum for different temperatures in the range 723 K to 1046 K are shown by the curves in Fig. 11.11. Here the wavelength in microns is plotted along the abscissa and the emissive power E_λ along the ordinate.

A study of the curves clearly indicates two facts. The first indication is that E_λ increases with temperature for all wavelengths. The second indication is that each curve has a definite maximum, (E_m, λ_m) and λ_m shifts towards smaller wavelengths as temperature increases.

The above findings support some interesting experimental observations. They justify why when a body is being heated its colour changes from red to white. They also support Stefan's law. The area enclosed between the curve and the abscissa represents the total radiation emitted by the body at a particular temperature T. By making necessary measurements it can be

seen that the area varies at T^4, and this is Stefan's law.

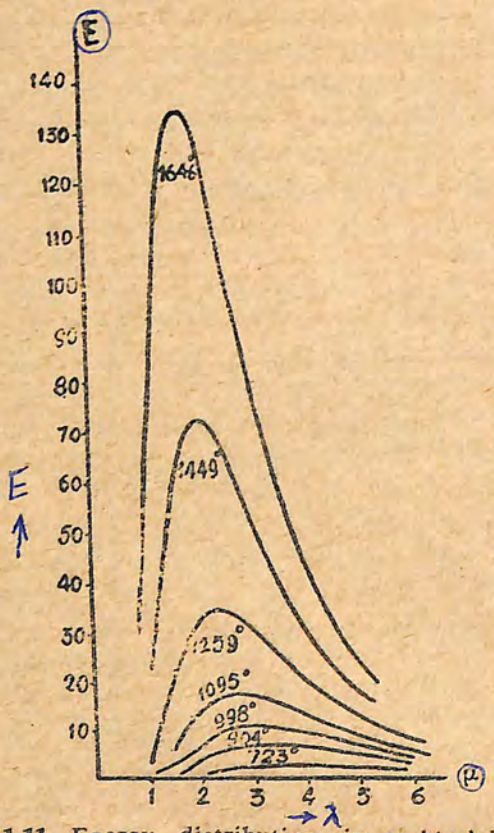


Fig. 11.11 Energy distribution in a black-body spectrum

11.22 Wien's Displacements Law

It was observed from Fig. 11.11 that λ_m shifts towards lower wavelengths as the temperature is increased. This observation can be quantitatively expressed by the equation.

$$\lambda_m T = \text{constant} \qquad ..(11.13)$$

Wien derived this equation on the basis of thermodynamic considerations and hence this is known as Wien's displacement law.

The constant is found to have the value

28.84×10^{-6} m-deg.

Wien's law can be used to determine the temperature of the sun and stars. For example, investigations have shown that λ_{max} for the sun is 4753 A°. On substituting this value in Wien's law the temperature of the sun comes out to be 6050 K. This value is slightly higher than the currently accepted value and this is because of the inherent assumption while using Wien's law that sun is a black body.

11.23 Pyrometers

Laws of radiations such as Stefan's law and Wien's law can be used to estimate temperature of hot bodies. But the temperature thus calculated will only be approximate because these laws are applicable strictly to black bodies and the actual hot bodies are seldom black.

Instruments developed to measure high temperatures using emitted radiation are called pyrometers. These differ from other instruments such as thermometers, thermocouples, etc. in that they are not to be in contact with the hot body. They can measure temperature however high and the lower practical limit for radiation pyrometers is about 900K. We will now briefly describe how to measure high temperatures using a pyrometer.

A typical radiation pyrometer is shown in Fig. 11.12. The beam of radiation from the hot

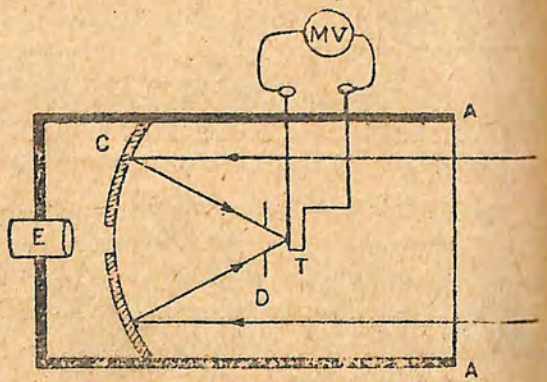


Fig. 11.12 A radiation pyrometer

body enters the instrument through the aperture AA. The concave mirror converges the beam. This focussed beam is made to pass through a limiting diaphragm D before it is incident on the thermocouple. The E.M.F. of the thermocouple is recorded on a millivoltmeter connected to the terminals of the thermocouple.

The E.M.F. developed is proportional to the temperature difference between the hot and the cold junctions. If T and T_0 are the temperatures of the hot body and the receiving junction, respectively, then the E.M.F. developed should be proportional to $T^4 - T_0^4$. Since T is very large as compared to T_0, T_0^4 can be neglected in comparision to T^4.

In practice it is found that E.M.F. is not exactly proportional to the fourth power of the absolute temperature of the source. Some of the reasons for this discrepancy are: (i) the source is not a black body, (ii) T_0^4 is not zero, and (iii) stray radiations and conduction of heat raise the temperature of the cold junction. So, before use, the pyrometer has to be suitably calibrated, and it is found that the value of the power varies from 3.8 to 4.2.

Exercises

11.1 Distinguish clearly between temperature and heat.

11.2 In a joule experiment two weights of 5 kg each fall through a height of 3m and rotate a paddle wheel which stirs 0.1 kg water. What is the change in the temperature of the water ? (0.7K)

11.3 A lead bullet weighing 20g and moving with a velocity of 100m/s comes to rest in a fixed block of wood. Calculate the heat developed and the rise in temperature of the bullet assuming that half the heat is absorbed by the bullet. The specific heat of lead is 0.03. (23.8 cal, 19.9K)

11.4 1g of water at 373K is converted into steam at the same temperature. The volume of 1 cc of water becomes 1671 cc on boiling. Calculate the change in the internal energy of the system if the heat of vaporisation is 540 cal/g. (499 cal)

11.5 Apply the first law of thermodynamics to obtain an expression for the change in the internal energy during the melting process.

11.6 Discuss whether the following phenomena are reversible :

A. Waterfall
B. Rusting of iron
C. Electrolysis.

11.7 Give two examples of a reversible process. Discuss their reversibility.

11.8 What is meant by a reversible engine ? Explain why the efficiency of a reversible engine is the maximum.

11.9 A steam engine takes steam from the boiler at 500 K and rejects it to the air at 373K. What is its efficiency ? (25.4%)

11.10 In a refrigerator heat from inside at 277K is transferred to a room at 300K. How many joules of heat will be delivered to the room for each joule of electric energy consumed ideally ? (13 joules)

11.11 What is meant by the emissive power of a surface ? Explain why the radiation in a cavity depends only on the temperature of the walls and not on the material of which it is made ?

11.12 Explain how Kirchoff's law helps to identify the elements in the sun's atmosphere. Why does a piece of red glass, when heated and taken out, glow with a green light ?

11.13 Wavelength corresponding to E_{max} for the moon is 14 microns. Estimate the temperature of the moon by taking the value of σ given in the book. (200K)

11.14 Describe a method to estimate the temperature of a hot furnace.

11.15 Using Stefan's law deduce Newton's law of cooling which states that "If the temperature difference between a hot body and its surroundings is not large then the rate of cooling is proportional to the excess of temperature of the substance above the surroundings."

UNIT 12

Liquids

12.1 Intermolecular Interactions

In kinetic theory of gases, we have seen that matter is composed of minute particles called molecules. It is assumed that there are no forces acting between them. In other words, the intermolecular interactions are supposed not to exist. But, in fact, these are simplifying assumptions. Even for gases, molecules have finite size and they interact with one another. It is for these reasons that the equation of state for a real gas differs from that for the ideal gas $PV=RT$. Van der Waals modified the equation of state of an ideal gas by making a correction in the pressure*. The reason for this correction is the presence of intermolecular interaction in real gases. This intermolecular interaction is therefore known as Van der Waals' interaction. It exists not only in gases but also in liquids and even in some solids (Art. 16.6). There are three general categories into which Van der Waals' interaction can be divided :

(1) Interaction due to dipole-dipole forces.
(2) Interaction due to induced dipole forces (also called induction forces).
(3) Interaction due to dispersion forces.

Van der Waals' interactions (each one of the above three) vary as $\frac{1}{r^6}$ where r is the intermolecular separation between the molecules. The dipole-dipole interaction is the strongest of the three and the dispersion interaction the weakest. Evidently the respective forces vary as $\frac{1}{r^7}$ with intermolecular separation.

Before proceeding further it will be worthwhile to have a look at the origin of the interactions. Even though the net charge on the molecule is zero, it is composed of positive and negative charges. In spite of the electrical neutrality of the molecule, however, the distribution of positive and negative charges on it is not even. The atoms in the molecule are so arranged that the centre of mass of the positive charges does not fall exactly on the centre of mass of the negative charges.

The positive and negative charges, separated by a distance, form an electric dipole. The molecular dipole has a dipole moment whose behaviour in an electric field is the same as that of a magnet in a magnetic field. Molecules having permanent dipole moment, like H_2O, are known as polar molecules, in contrast

* Van der Waals also applied a correction in volume due to the finite size of the molecules. We are here not concerned with this aspect.

to non-polar molecules like CO_2, O_2, N_2, etc. Since the dipole-dipole interaction is strong, generally polar molecules are liquid under ordinary physical conditions (water, alcohol, etc.).

The nature of induction and dispersion forces is difficult to understand and will not be treated at this stage. However, the ultimate nature of induction as well as dispersion forces is also of the dipole-dipole type. The dispersion force, which is the weakest of the three, can occur in all cases and even in monoatomic gases like argon, etc.

Thus, it is seen that the molecular interactions are electric in nature. The molecular forces are (i) attractive at large distances, and (ii) have necessarily to be repulsive at very close distances. The nature of intermolecular forces has been shown in Fig 12.1. The attrac-

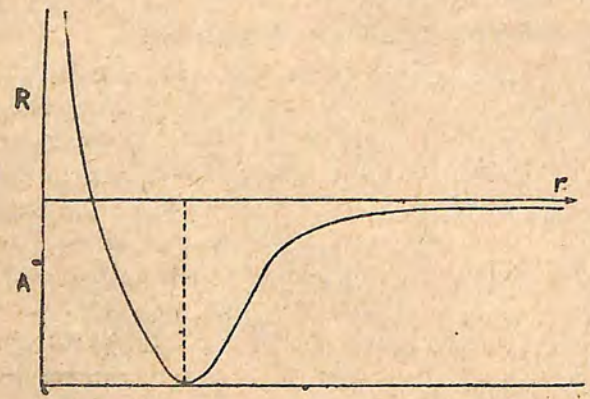


Fig. 12.1 Nature of intermolecular force
R—repulsive force. A—Attractive force.
r—separation

tive force increases rapidly with decrease in separation between molecules for larger distances. Some time it is convenient to imagine a small sphere of radius of about 10^{-9}m around a molecule and call it its sphere of action. Since interactions fall rapidly with distance it is useful to regard the molecule as exerting an attractive force on other molecules lying within its sphere of action and neglect its influence on others lying outside this sphere. In some diagrams in this unit this sphere has been shown by a dashed circle.

When the separation becomes too small the repulsive forces start dominating even more rapidly. Had there been no repulsive forces all matter would have collapsed to a point. The intermolecular separation (r_e) maintains an equilibrium position where the potential energy due to repulsion and attraction is a minimum. These repulsive forces are also electric in origin and are due to point charges in the atoms of the molecule which form the molecule.

The minimum potential energy corresponds to $r_e \simeq 10^{-9}$ m and the repulsive force varies more sharply with decreasing intermolecular distance, approximately as $\frac{1}{r^9}$.

The combination of the intermolecular forces and thermal motion gives rise to three states of matter, i.e. solid, liquid and gas. In solids the attractive forces are quite strong and thermal motions cannot break them away. Molecules may stay at one place and vibrate. In gases the intermolecular attractive forces are so weak that random thermal motion can very easily overcome them and molecules are almost free to move about anywhere. Liquids fall in between these two cases. For the sake of simplicity a liquid is sometimes visualized as a dense gas. The molecules in a liquid are neither forced to stay permanently in an equilibrium position nor they are free to leave the company of other molecules. A molecule inside a liquid can slide freely over others without any effort.

Cohesive and adhesive forces

The attractive forces between similar molecules discussed above are called cohesive forces.

The forces between dissimilar molecules e.g. the molecules of glass and water are called adhesive forces. The attractions are called cohesion and adhesion respectively. We can thus define a liquid as a state of matter in which the molecules can change their position with respect to each other but are restricted by molecular forces or cohesive forces so as to maintain a fixed volume. It is also known that if we want to convert a liquid into vapour, we have to supply an extra amount of energy. To separate two molecules the energy supplied to them should at least be equal to the intermolecular interaction energy. Intermolecular energy in liquid molecules can be calculated from the latent heat of vaporization of the liquid.

EXAMPLE 12.1

The latent heat of vaporization for water is 22.6×10^5 Joule/kg. Calculate the intermolecular binding energy. (Avogadro Number $N=6\times10^{23}$ mole^{-1}, 1 Joule$=0.62\times10^{19}$eV).

Solution

Molecular weight of water$=18$

Number of molecules in 1kg of water

$N=6\times10^{26}/18=10^{26}/3$.

Energy required to unbound

N molecules of water$=22.6\times10^{+5}\times0.62\times10^{19}$eV

$=1.4\times10^{25}$ eV.

Intermolecular binding energy

$$=\frac{1.4\times10^{25}\times3}{10^{26}}\text{ eV}$$

$=0.4$eV.

12.2 Surface Tension

A dead fly or a beetle tends to sink to the bottom when immersed in water. This indicates that its average density is larger than that of water. And yet we see in nature large water beetles running on the surface of a lake seemingly without wetting their feet. A sewing needle placed carefully on the surface of a liquid floats at the surface, though the density of the needle's materials may be as much as eight times that of water. A liquid flowing slowly from the tip of a medicine dropper does not emerge as a continuous stream but as a succession of drops. Rain drops, fog drops, soap bubbles, etc. assume spherical shapes as they fall through the air. All these phenomena, and many others of similar nature, are associated with the existence of a boundary surface between a liquid and some other substance. Let us now consider the properties of the liquid surface.

From the consideration of a molecular interaction we have seen that unlike solids, liquids do not have any definite shape of their own. Even though weakly, the molecules of a liquid interact with one another. Liquids attain the shape of the container in which they are kept. Due to gravitation the surface of a liquid is plane and horizontal in a wide vessel. In the presence of adhesive forces, that is, interaction of the molecules of the material of container and molecules of the liquid the shape of the liquid surface is deformed near the walls of the container. In narrow tubes, therefore, the surface of the liquid appears concave or convex. However, in wide vessels, as stated above, the distortion which is only at the boundary is negligible and the surface of the liquid is plane. The intermolecular (or cohesive) forces discussed in Art. 12.1 explain the interesting behaviour of a liquid (Fig. 12.2 a). It is surrounded by similar molecules from all the sides. The net force on such a molecule is zero. All the molecules inside the liquid are in similar approximate environment and have more or less the same potential energy. Hence, these molecules slide from one place to another inside the liquid without doing any work. The situation is different when we consider a molecule. B on the surface

surface which is attracted by many molecules from below but there is no attractive force from above the surface (the force due to molecules

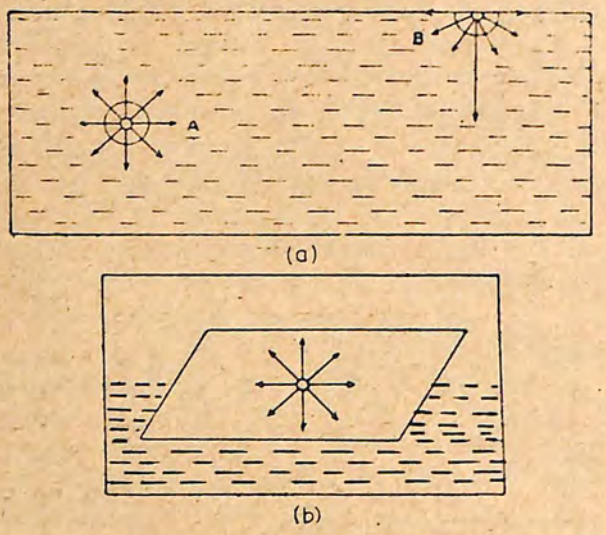


Fig. 12.2 (a) Molecule inside a liquid
(b) Molecule on the surface of a liquid

in liquid vapour can be neglected). A molecule on the surface will therefore be attracted in all directions by the molecules beneath. We can resolve all these forces in the vertically downward direction. The resultant picture is that every molecule on the surface is acted on by a vertically downward pull inside the liquid. Due to this downward force, the potential energy of the molecules on the surface is larger than that of those inside. In consequence, the surface of a liquid tends to contract as much as possible. In other words, the surface of a liquid is under tension. In Fig. 12.2 (b) a molecule on the surface is shown to be acted upon by the molecular forces of the neighbouring surface molecules symmetrically in all directions. The net force is zero but the surface acts like a stretched membrane. Such a tension is measured in a stretched membrane across unit length of an imaginary straight line drawn on the surface. The surface tension of a liquid therefore is defined as the force of contraction across an imaginary line of unit length tangential to the surface of a liquid, the line and the force being perpendicular to each other. The unit of surface tension is dyne per cm in c.g.s. or Newton per metre (Nm^{-1}) in MKS, system.

TABLE 12.1

Surface Tension of Some Liquids

Liquid	*In Contact with*	*Newton/metre* $\times 10^{-3}$
Olive oil	air	35
Benzene	air	29
Glycerine	air	63
Water	air	75
Mercury	air	35

As an example we see that a bug floats on water due to surface tension. The bug (Fig. 12.3) bends its legs on the surface of water such that the deformed water surface gives rise to forces of surface tension along the directions as shown in Fig. 12.3. The weight of the bug

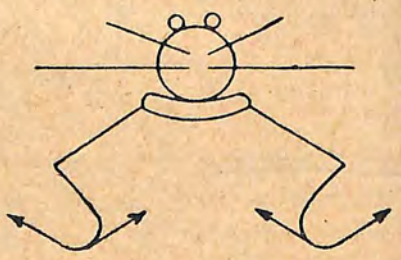


Fig. 12.3 A bug upheld on water surface by surface tension

is neutralized not by buoyancy (the bug is hardly submerged) but by the upward components of the surface tension forces which are tangential to the deformed surfaces.

As mentioned earlier, another example to understand the phenomenon of surface tension is that liquids show a tendency to assume a characteristic spherical shape. A molecule of a liquid deep inside a drop is surrounded from all sides by other molecules. To remove this molecule A (Fig. 12.4) from inside the

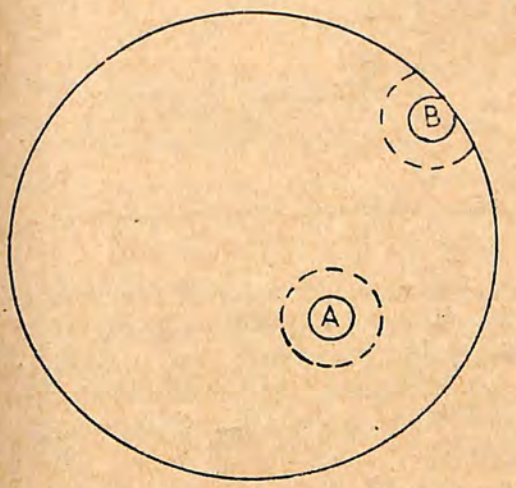


Fig. 12.4 A—molecule inside a drop
B—molecule on the surface of the drop

drop we have to do work against the attraction of all its neighbours. Since we must do some work to remove this molecule from the drop, it apparently had quite a large negative energy, say $-E_1$, when it was inside the drop. Let us repeat the process for a molecule B, on the outer surface of the drop. This molecule, being on the surface, is not surrounded by neighbours on all sides except from beneath. Therefore, less work will have to be done to remove this molecule. Let the energy of the molecule B on the surface be $-E_s$. It is obvious that $|E_s| < |E_1|$ or $-E_s > -E_1$, that is, the energy at the surface is positive as compared to the energy in the interior of the drop. We know that the physical systems come to a state of least energy if they are able to (a car rolls down hill, a hot body becomes cooler, and so on) reduce the number of its high energy molecules and therefore its total energy. So a piece of liquid will do its best to make its surface as small as possible. The geometrical shape that has the least surface for a given volume is a sphere; so the liquid comes as close as it can to assume the form of a spherical drop. The departure from the spherical shape is due to the weight of the drop as well as the adhesive forces if any. It is seen that as soon as the water drop falls on a glass plate it flattens. This is due to the adhesive force between glass molecules and water molecules.

12.3 Surface Energy

We have seen that the molecules on the surface of a liquid which are pulled inward (like weights on a shelf which are pulled earthward by gravity) possess potential energy. As a molecule enters the interior it loses energy. Thus, because a molecule of a liquid on the surface is at higher potential energy than a molecule of the liquid, deep inside, one has to do certain amount of work in creating a liquid surface. In other words, if we want to take a molecule from inside the liquid to the surface, we will have to work against the downward pull. Thus, if we want to create a surface, the work of increasing the number of molecules on the surface has to be done. The energy required to create a surface is known as surface energy. We can calculate this energy by means of a simple experiment. A rectangular frame of wire, ABCD, is taken. The arm AD of the frame is not fixed to the frame but can slide,

resting on the arms CD and BA. The frame is carefully dipped in a soap solution so that a film ABCD is formed. Let the sliding arm AD be of length l (Fig. 12.5) and the surface tension

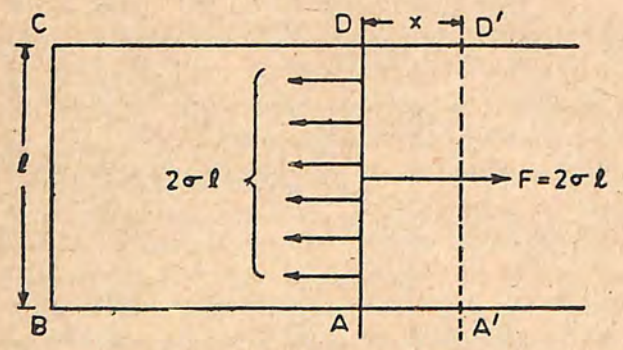


Fig. 12.5 Increasing the area of the drop

of the liquid be σ. Then by definition of surface tension, the force acting on the wire is

$F = \sigma 2l$, since the film has two surface.

If DA is slid to a new position D′A′ by a distance x, then the increase in the surface area of the film is

$\triangle S = 2lx$

The work done in creating this new surface $\triangle S$ is given by

$$W = F\,x$$
$$= \sigma\, 2lx$$
$$= \sigma\, \triangle S$$

$= \sigma \times$ increase in the area of the film.

Thus, the work done to increase the surface area of a film by unit amount is equal to the surface tension of the liquid. Therefore, σ is also called the free surface energy per unit area. In actual practice the force acting on wire AD (Fig. 12.6) can be measured by a spring balance. Suspend a frame of wire like ABCD from the hook of the balance. CD is fixed to the frame and is not free to slide. As

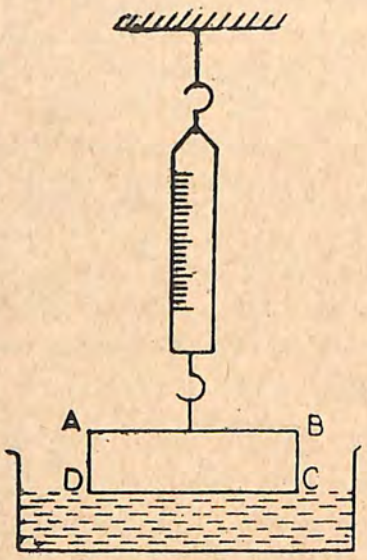


Fig. 12.6 Measuring the force due to surface tension

soon as the arm CD is brought in contact with the surface of a liquid in a beaker, the wire is pulled down. This force can be measured by the spring balance. It will be found that the force Fl which is the pull on the spring, is equal to $2\sigma l$.

EXAMPLE 12.2

When two mercury drops are brought into contact, they form one drop. Explain.

Solution

Let each of the small drops have a radius r,

$$\text{Volume of each drop} = \frac{4}{3}\pi r^3$$

Surface area of each drop $= 4\pi r^2$

Total surface energy of the system of two drops $= 8\pi r^2 \sigma$

where σ = surface tension of mercury.

When both drops unite, the radius of the resulting drop is given by the equation

$$\frac{4}{3}\pi R^3 = \frac{8}{3}\pi r^3$$

where R is the radius of the resulting drop.

So $R = 2^{\frac{1}{3}} r$ and

Surface energy of single larger drop

$= 4\pi r^2 2^{\frac{2}{3}}$.

Hence the surface energy of the large drop is less than the total surface energy of the smaller drops. As every physical system tries to attain a state of minimum energy, the drops of mercury coming in contact form one drop.

Conversely, if you want to break any drop of liquid into smaller drops you will have to do work.

The low value of surface tension helps the liquid to spread out as a thin film as the spreading is easier on account of low surface energy. Low surface tension also allows a liquid to penetrate in narrow spaces or pores. Water has relatively high surface tension. But if detergent is added to it, the surface tension is lowered. The cleaning property of the solution is improved because now the solution can seep into fine pores where water could not.

So far we have considered only a free liquid surface. But actually a liquid surface is either in contact with a solid or vapour or some other liquid. Consider a system consisting of a solid S and liquid L (Fig. 12.7). Suppose

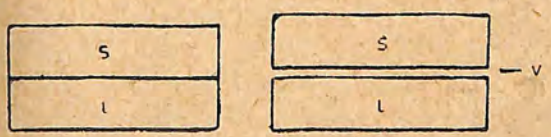


Fig. 12.7 A system containing a solid S, liquid L and vapour V

that they are initially in contact, and are then separated. Let the work per unit area required to separate them be W_{SL}. Before separation there is potential energy in the interface amounting to σ_{SL} per unit area (σ_{SL} is the surface tension of solid-liquid film). After separation there is energy σ_{SV} per unit area on the surface between the solid and vapour (σ_{SV} is the surface tension of solid-vapour film) and σ_{LV} per unit area on the surface between liquid and the vapour (σ_{LV} is the surface tension of liquid-vapour film). The initial energy of the system plus the work done on the system for separating the solid from liquid is equal to the final energy after separation. Hence we have

$$\sigma_{SL} + W_{SL} = \sigma_{SV} + \sigma_{LV} \qquad \ldots (12.1)$$

12.4 Angle of Contact

Liquid has to be kept in a container. The molecules of container are different from those of liquid. On the walls of the container three surface films viz, liquid-vapour, solid-vapour and solid-liquid are formed. The films are only a few molecules thick. Associated with each film is an appropriate surface tension, (Fig. 12.8). The curvature of the surface of liquid near a solid wall (or the shape of a liquid drop on a surface) depends upon the difference between σ_{SV} and σ_{SL}. At the wall the three films meet. If we isolate a small portion of all three films at their junction and imagine the film to extend a unit distance perpendicular to the diagram, the isolated portion will be in equilibrium under the action of four forces, three of which are surface tensions of the three films. The fourth force F_a is adhesive force between isolated portion of film and the wall. From equilibrium condition, we get

$$\sigma_{LV} \sin\theta = F_a$$

$$\sigma_{LV} \cos\theta = \sigma_{SV} - \sigma_{SL}$$

θ is the angle between σ_{LV} and σ_{SL} taken inside the liquid and it is called the angle of

contact. Eliminating

$\sigma_{SV} - \sigma_{SL}$ we get

$$W_{SL} = \sigma_{LV}(1 + \cos\theta) \quad \ldots (12.2)$$

If the work done per unit area required to separate liquid from the solid is greater than

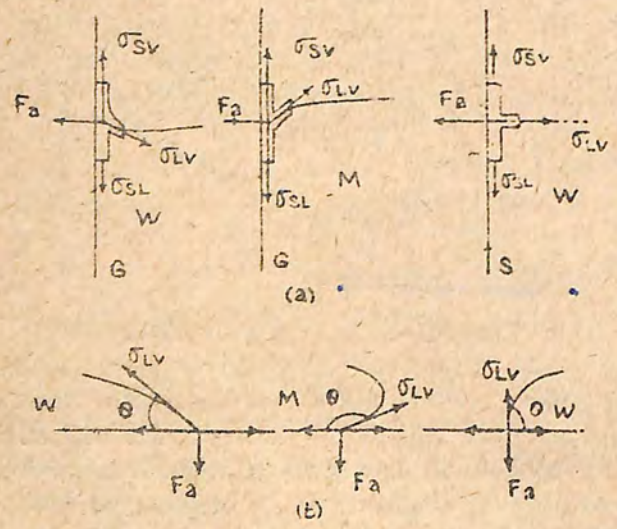


Fig. 12.8 (a) Layer of a liquid near the wall of a container
W—water
M—mercury, S—silver, G—glass
(b) Drop of a liquid on a surface

the surface tension of liquid-vapour film the angle of contact θ is acute, that when $W_{SL} > \sigma_{SV}$, $\theta < 90°$ and $W_{SL} < \sigma_{LV}$ $\theta > 90°$.

When water is taken in a thin glass tube the meniscus of water is concave, where as the meniscus of mercury is convex. This is because the work done in removing a unit area of water surface from the glass surface is more than the surface tension of water-vapour film. Another way of explaining it is that in the case of water-glass-water vapour system we have $\sigma_{SV} > \sigma_{SL}$. In case of mercury-glass-mercury vapour system $\sigma_{SV} < \sigma_{SL}$, the angle of contact is obtuse. Thus the meniscus of mercury in glass is convex. When the water is kept in a silver container $\sigma_{SV} \cong \sigma_{SL}$ and θ is 90°.

A solid surface will be wetted by the liquid if θ is small, but when θ is larger than 90° the liquid will slip over the surface. Impurities and adulterants present in or added to a liquid may alter the angle of contact considerably. Water proofing agents applied to a cloth cause the contact angle of water in contact with the cloth to be larger than 90°.

12.5 Pressure Difference Across a Surface Film

When the free surface of a liquid is plane (Fig. 12.9 (a) the resultant force due to surface tension on a molecule in the surface is zero. If the surface is curved, then there is a resultant force normal to the surface. This resultant force is directed into the liquid in the case of convex surface (Fig. 12.9 (b). In order to keep

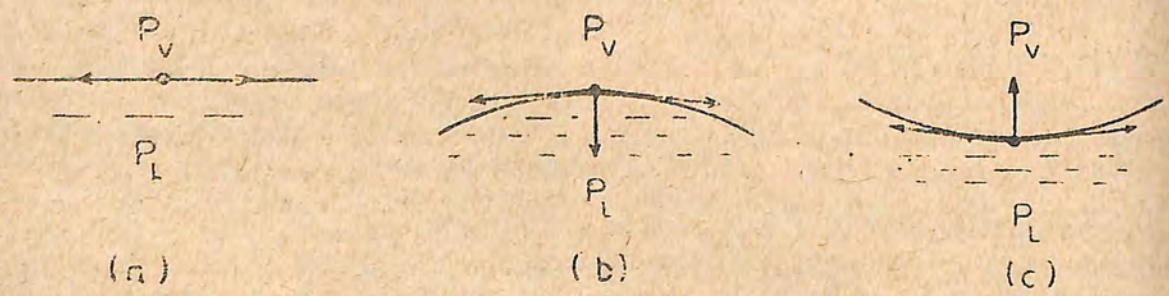


Fig. 12.9 Pressure difference across a surface film :
(a) Plane surface; (b) convex surface; (c) concave surface.

this surface in equilibrium the pressure on liquid side of the surface must be more than the pressure on the vapour side, is, $P_L > P_v$. Similarly, it can be shown in the case of a concave surface $P_v > P_L$. Hence, we conclude that the pressure on the concave side of the film is always greater than the pressure on the convex side.

The argument given already shows that whenever the surface of a liquid is curved there exists a difference of pressure inside and outside the surface. Consider for example a spherical drop of a liquid of radius R. Due to the spherical shape there will exist excess pressure inside the liquid as compared to outside atmospheric pressure. Let the excess be P. Let us enlarge the liquid drop from radius R to R+dR. This increases the surface area of the drop from S to S+dS.

Since $S = 4\pi R^2$ $\quad dS = 8\pi R dR$

From the definition of surface tension the work done in increasing this surface area is σdS. Since work done is also equal to force multiplied by distance and force is equal to pressure × area. We have

$$P \times 4\pi R^2 \, dR = \sigma 8\pi R dR$$

$$\text{or } P = \frac{2\sigma}{R} \qquad \ldots (12.3)$$

It follows from this result that if the surface tension remains constant, the pressure difference is larger for smaller value of R. Due to this large internal pressure, tiny fog drops have rigidity properties like those of solids. A good illustration is to be found in the case with which ice-skates slide over the surface of smooth ice. Under the enormous pressure exerted on the ice by the sharp metal edge of the skate, ice melts, and the runners run along on the tiny drops as if they were on ball bearings.

EXAMPLE

Find the difference in air pressure between inside and outside of a soap bubble 5 mm in diameter. Assume the surface tension to be 1.6 N m^{-1}.

Solution

Unlike a drop of liquid, the soap bubble with air inside has two surfaces, one external and another internal. Due to two spherical surfaces the expression for the excess of pressure inside a soap bubble is given by

$$P_{excess} = 4\sigma/R,$$

$$= \frac{4 \times 1.6}{2.5 \times 10^{-3}} = 2560 \text{ N m}^{-2}$$

12.6 Capillarity

The most familiar surface effect is the elevation of a liquid in an open tube of small cross-section. The term 'Capillarity', used to describe effects of this sort, originates from the description of such tubes as capillary or 'hair-like' (Fig. 12.10). In case of a liquid that wets the tube, the contact angle is less than 90° and the liquid rises until an equilibrium height 'h' is reached. When the contact angle is greater than 90° the liquid is depressed inside the capillary as in mercury. If the pressure on the concave side of the meniscus is P then that on the convex side is $P' = P - \frac{2\sigma}{R}$ (from equation (12.3) since we can very easily see that P—P′=the pressure due to the liquid column of height 'h').

In case of a liquid forming a concave meniscus in a glass capillary : P is the atmospheric pressure, and as $P' < P$ the liquid must rise in the capillary till the weight of the water column is such that it exerts a pressure P—P′ at its base. If the angle of contact is θ, and the radius of capillary is R, the radius of meniscus

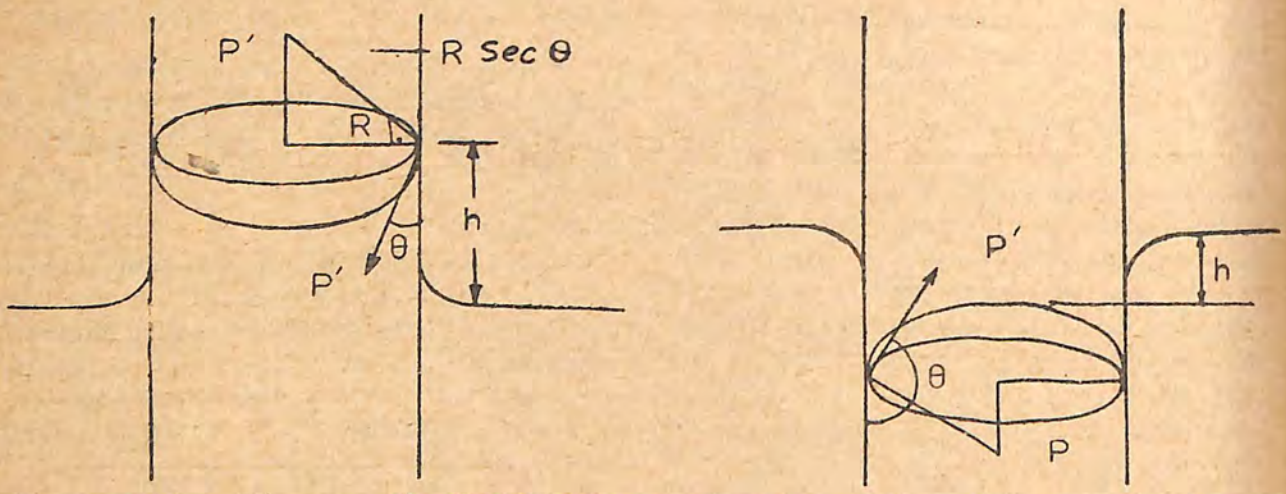


Fig. 12.10 Liquid column inside a capillary

is given by R secθ. The excess of pressure is given by

$P - P' = 2\sigma/R \sec\theta$

Therefore $2\sigma \cos\theta/R = h\rho g$ (ρ = density of the iquid)

$$\text{or } h = \frac{2\sigma \cos\theta}{R\rho g} \quad \ldots (12.4)$$

Here we have neglected the small volume of he meniscus. In case of water, $\theta = 0$,

$$h = \frac{2\sigma}{R\rho g}$$

The same equations hold for capillary depression.

We can find out the surface tension of a liquid if we know the angle of contact. The rise of liquid in capillary and the radius of capillary are measured with the help of a travelling microscope. It should be noted that before inserting the tube in the liquid, the tube must be cleaned. If there is any impurity at the inside surface of the capillary, the angle of contact will change, thus causing error in the measurement. The capillarity phenomenon is responsible for the rise of oil in a wick lamp, the absorption of ink in a blotting paper, etc. If the height of capillary is smaller than h given by equation (12.4) the water will not continue to flow out of the top of capillary as a fountain. The uppermost periphery of water will not be in contact with the walls of capillary. The actual shape of meniscus will change and the radius of curvature of meniscus will be such that the difference of pressure across the meniscus will only maintain a water column of smaller height.

12.7 Flow of Liquids

In summer the flow of water in rivers and canals seems to be steady. The same rivers when in flood show a very unsteady flow. There appear whirl pools and vortices on water surface. If a liquid in a vessel is stirred and then left to itself, the motion will disappear after some time. These examples show that liquids under different conditions flow differently. In flowing liquids, it is not necessary, that at any particular instant all the molecules have the same velocity. The molecules or molecular layers can move relative to each other with different velocities. Due to relative motion between different molecular layers, frictional force acts, causing resistance to the motion of the liquid. This resistive

property is known as viscosity. Therefore, in order to slide one layer of liquid over another an external force must be exerted to provide uniform motion to a liquid.

Let now the particles of a liquid be in motion. When the liquid velocity at any given point is constant of time, the motion is said to be steady. Whichever particle passes through the point A, it has a velocity V_A in a steady flow (Fig. 12.11). When one particle leaves A

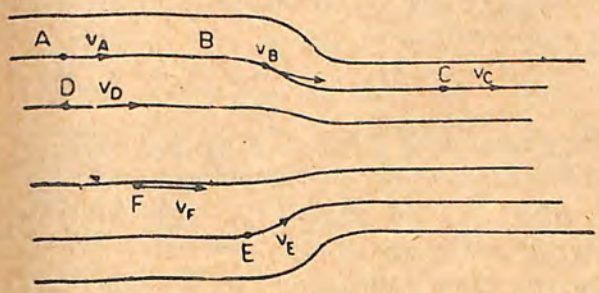


Fig. 12.11 Steady flow of a liquid

with velocity V_A, it is followed by another particle. When this second particle leaves A it also has a velocity V_A. Similar thing must occur at all other points in the liquid, that is, the liquid particle must always leave B with velocity V_B, C with velocity V_C, D with velocity V_D and so on. It is not necessary for these velocities to be equal.

In a steady flow let a liquid particle follow the path ABC. If the liquid particles preceding it and following it also follow the same path as ABC, the path is called a line of flow or a stream-line. A stream-line is parallel to the velocity of fluid particles. No two stream-lines can cross each other. The stream-line motion is possible only when the liquid velocity is low and the velocity along a path does not change too abruptly. If the velocity is high, or the change in velocity is frequent and very abrupt, the motion ceases to be a stream-line flow.

In non-viscous liquids, the velocity of all the particles at one section of a pipe are equal and liquid advances as a unit along a pipe. The surface determined by heads of velocity vectors is a plane, and the fluid flow is characterized by a plane velocity profile (Fig. 12.12 a). When the fluid is viscous and the velocity is not large, the flow is called laminar. (The lamina of liquid slide over each other). The velocity profile has the shape shown in Fig. 12.12. The velocity is maximum along

Fig. 12.12 Flow of liquid through a pipe
(a) Non-viscous liquid
(b) viscous liquid

the axis of the pipe and decreases to zero at the wall. When the velocity exceeds a certain critical value, the nature of flow becomes very much complicated. Random, irregular, local circular currents called vortices, develop throughout the fluid. The resistance to the flow increases tremendously. This type of flow is called turbulent flow.

It has been observed experimentally that there is a combination of four factors which determines the nature of flow of a viscous fluid through a pipe. This combination is known as the *Reynold number*, N_R and it is defind as

$$N_R = \frac{\rho v D}{\eta} \qquad \ldots (12.5)$$

Here ρ is the density of the liquid, v is its average speed, D is the diameter of the pipe and η is the coefficient of viscosity of the

liquid. The flow of viscous fluid is said to be laminar when N_R lies between 0 and 2000. For values of N_R above about 3000, the flow is turbulent. For N_R between 2000 and 3000 the flow is unstable and may switch over from one type to another. Reynold number is a pure number and therefore its numerical value is the same in any consistent set of units.

EXAMPLE 12.4

What should be the maximum average velocity of water in a tube of diameter 2.5 cm so that the flow is laminar ? The viscosity of water is 0.001 $Nm^{-2}S$.

Solution

The maximum value of Reynold number for laminar flow is 2000. Let V be the maximum average velocity of water for laminar flow. Then

$$\frac{\rho VD}{\eta} = 2000$$

$$\text{or} \quad V = \frac{2000 \times 0.001}{10^3 \times 2.5 \times 10^{-2}}$$

$$= 0.08 \text{ ms}^{-1}$$

12.8 Viscosity

As mentioned earlier viscosity may be thought of as the internal friction of a liquid. Because of viscosity, a force must be exerted to cause one layer of liquid to slide past another. If there is a layer of liquid between two sheets a force is needed to slide one sheet over another. Let a liquid flow over a fixed surface such as AB (Fig. 12.13). The layer of liquid in contact with AB remains at rest while the upper most layer PQ has the maximum velocity V. The laver EF at a distance x + dx from AB flows with a greater velocity than that of a layer CD at a distance x from AB. If the difference in the velocities of these two layers is dV, the velocity gradient between layers CD and EF will be dV/dx. Every lower layer tries to check the flow of the adjascent upper layer. This is eqiuvalent to the statement that the flow of the liquid is opposed by

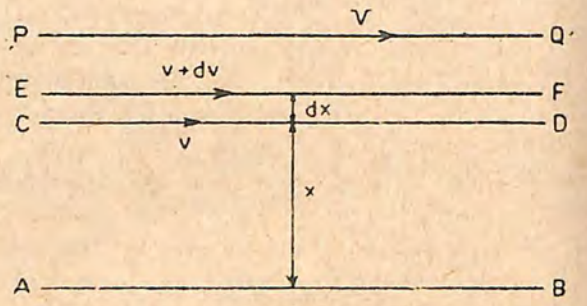


Fig. 12.13 Flow of viscous liquid over a surface

a tangential force F offered along the surface of the layer. It is found by experiment that the force per unit area offered by liquid layers is proportional to the velocity gradient. That is,

$$F/A \propto dV/dx$$

$$\text{or} \quad F = \eta A. \, dV/dx \qquad \ldots (12.6)$$

The proportionality constant η, is called the coefficient of viscosity or simply the viscosity. Poise is defined as a tangential force per unit area offered by a liquid layer to create a unit velocity gradient. The dimension of η can be easily calculated as

$$\eta = \frac{F}{A} \Big/ \frac{dV}{dx} = \frac{MLT^{-2}}{L^2} \Big/ \frac{L}{TL}$$

$$= ML^{-1}T^{-1}$$

We have already seen that when a viscous liquid flows through a tube, the flow is laminar depending upon the radius of the cross-section of the tube and the average velocity of liquid. The force required to maintain a laminar flow of liquid through a tube of length l and radius r is given by

$$F = 8\eta l \frac{V'}{r^2} \qquad \ldots (12.7)$$

where V' is the volume of the liquid flowing per second through the tube.

Stoke's Law

So far we have discussed the flow of a liquid on a solid surface or through a pipe. Let us now consider the flow of a viscous liquid past a body or the motion of a body through a viscous fluid at rest. A steel ball falling in a tank of glycerine attains a constant velocity in a very short distance from the start of the motion. This shows that the retardation due to resisting force on the steel ball is equal to the acceleration due to gravity. When a spherical body moves through a viscous liquid at rest (Fig. 12.14), it is expected that the

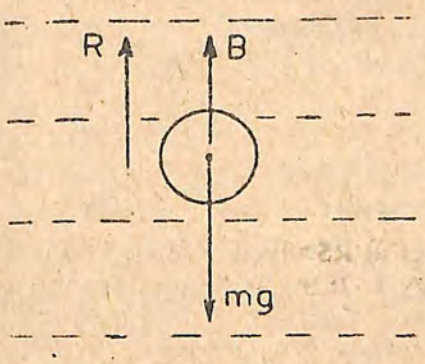


Fig. 12.14 Motion of a sphere through a fluid

resisting force will depend upon (i) the size of the body (ii) the relative speed of the body with respect to the liquid and the viscosity. Thus R, the resisting force on the body is given by

$$R = \text{constant } r^a v^b \eta^c$$

where a, b and c are respectively the powers of radius r, velocity v and viscosity η. From dimensional analysis we see that the dimensions on left hand side must be equal to the dimensions on the right hand side of the equation. That is,

$$MLT^{-2} = (L)^a (LT^{-1})^b (ML^{-1}T^{-1})^c$$
$$= (M)^c (L)^{a+b-c} (T)^{-b-c}$$

Hence $a = b = c = 1$. We get

$$R = \text{constant } \eta r v$$

The constant is actually equal to 6π and thus we can write the resisting force as

$$R = 6\pi \eta r v \quad \ldots (12.8)$$

This equation is called Stoke's law. We shall consider it briefly in relation to a sphere of density ρ falling through a viscous fluid of density ρ_o. The forces acting on the sphere are :

(i) Weight of the sphere (acting downward)

$$mg = \frac{4\pi}{3} \rho \quad r^3 g.$$

(ii) Force of buoyancy (acting vertically upward)

$$B = \frac{4\pi}{3} r^3 g \rho_o.$$

(iii) Resisting force of liquid

$$R = 6\pi\eta \, rv.$$

The resultant acceleration 'a' with which the sphere is falling downward is given by

$$ma = mg - B - R$$
$$a = g - \frac{(B+R)}{m}$$

If the sphere is released from rest ($v = 0$) the viscous force R at the start is zero. The initial acceleration is therefore

$$a = g \frac{\rho - \rho_o}{\rho}$$

As a result of this acceleration, the sphere acquires a downward velocity and therefore experiences a retarding force given by Stoke's law. With increase in velocity the viscous resistance increases. An ultimate constant velocity will be attained by the sphere when the downward acceleration is reduced to zero, and the sphere moves with a constant velocity called its terminal velocity.

Putting a=0 in equation, we get

$$\frac{4}{3}\pi r^3 \rho g = \frac{4\pi}{3} r^3 \rho_o g + 6\pi\eta r v$$

$$\text{or} \quad v = \frac{2}{9}\frac{r^2 g}{\eta}(\rho - \rho_o) \qquad \ldots (12.9)$$

Equation holds provided the velocity is not so great that turbulence sets in. In presence of turbulence the resistance offered to the motion of sphere is much greater than that given by Stoke's law.

EXAMPLE 12.5

Find the terminal velocity of a steel ball 2 mm in diameter, falling through glycerine. Sp. gr. of steel=8

Sp. gr. of glycerine=1.3
viscosity of glycerine=8.3 poise.
The terminal velocity is given by

$$v = \frac{2}{9} \times \frac{(.1)^2 \times 980}{8.3}(8.3-1.3)$$

$$= 1.8 \text{ cm sec}^{-1}$$

This method is one of the methods used for measuring viscosity.

12.9 Bernoulli's Equation

Every liquid has a viscosity. The equation of motion of a viscous fluid is very complicated. But it is seen that many physical phenomena are explained qualitatively and to a good approximation even when we neglect the effect of viscosity. Let us now consider an incompressible, non-viscous liquid having a steady flow through a pipe. Let the pipe have a uniform cross-section A_1 at PQ at a height h_1 and a uniform cross-section A_2 at RS at a height h_2 (Fig. 12.15). Let at A_1 the liquid pressure be p_1 and the liquid velocity be $\bar{v}_1$. In a small time interval $\triangle t$ the layer at PQ is pushed to P'Q' under the influence of a force p_1A_1. Thus the displacement of PQ is $l_1 = \bar{v}_1 \triangle t$.

Hence, the work done on the system is

$$W_1 = p_1 A_1 l_1$$
$$= p_1 A_1 \bar{v}_1 \triangle t$$

The layer at RS advances to R' S' by a distance $l_2 = v_2 \triangle t$. If the liquid pressure and velocity

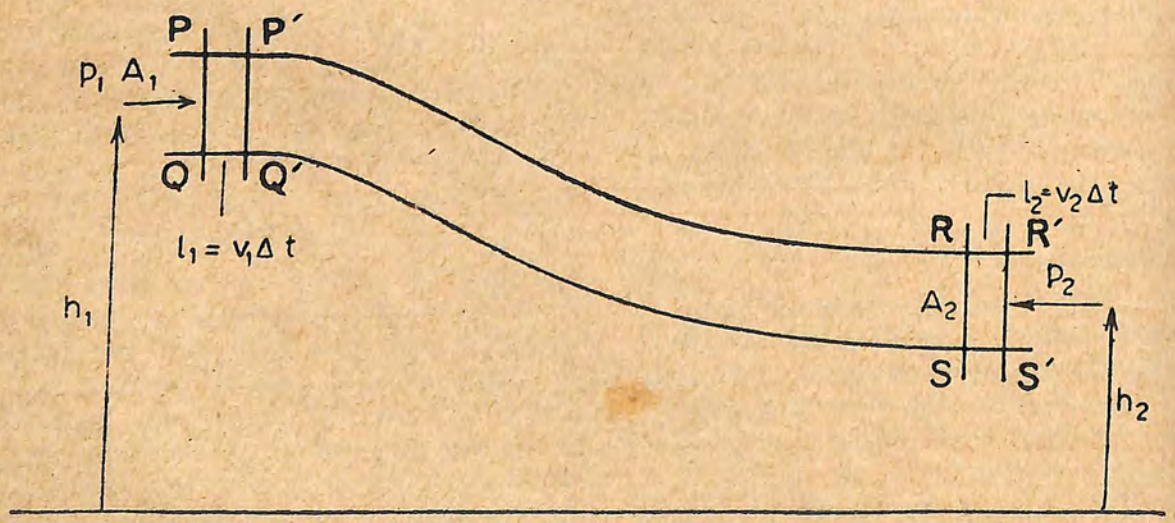


Fig. 12.15 Flow of liquid through a pipe with its ends at different heights

are respectively p_2 and v_2 at A_2, the work done by the system is :

$$W_2 = p_2 A_2 l_2$$
$$= p_2 A_2 v_2 \triangle t.$$

The net work done by the liquid pressure is

$$W_p = W_1 - W_2$$
$$= p_1 A_1 l_1 - P_2 A_2 l_2$$
$$= (p_1 - p_2) V$$

Hence $V (= A_1 l_1 = A_2 l_2)$, is the volume of the liquid that flows through the pipe in time $\triangle t$. From the above we see that

$$V = A_1 v_1 \triangle t = A_2 v_2 \triangle t$$
$$\text{i.e. } A_1 v_1 = A_2 v_2 \quad \ldots (12.10)$$

Equation 12.10 is called the equation of continuity. It shows that the speed of liquid flow is inversely proportional to the area of cross-section of the pipe.

During the time $\triangle t$, the height of the mass ρV changes from h_1 to h_2. The work done on the liquid by gravitational field is

$$W_g = -(h_2 - h_1) \rho v g$$
$$= \rho g V (h_1 - h_2)$$

We know that when work is done on an element of mass ρV, its kinetic energy changes. Hence

$$\text{change in K.E.} = W_p + W_g$$
$$= \text{Total work done on the system}$$

$$\text{or } \frac{1}{2}\rho V (v_2^2 - v_1^2) = (p_1 - p_2) V + \rho g V (h_1 - h_2)$$

$$\text{or } \frac{p_1}{\rho g} + \frac{v_1^2}{2g} + h_1 = \frac{p_2}{\rho g} + \frac{v_2^2}{2g} + h_2$$

$$\text{or } \boxed{\frac{p}{\rho g} + \frac{v^2}{2g} + h = \text{constant}} \quad \ldots (12.11)$$

Equation 12.11 is Bernoulli's equation. Every term in this equation has dimensions of length and is called a head. The $p/\rho g$ is called pressure head, $v^2/2g$ the velocity head and h the gravitational head. We may state, 'where the velocity is high the pressure is low, and where the velocity of a fluid is low the pressure is high.' We can discuss some illustrative examples of Bernoulli's principle.

(i) During certain wind storm or cyclone the roof of some houses are blown off without damaging the other parts of house (Fig. 12.16). This is not a freakish accident as one might think, for there is a simple explanation: A high wind blowing over the roof, creates a low pressure p_2 on top. Under the roof the pressure is p_1 (=atmospheric pressure) which is larger than p_2. The difference of pressure $(p_1 - p_2)$ causes an upward thrust and the roof is lifted up. Once the roof is lifted it is blown off with the wind.

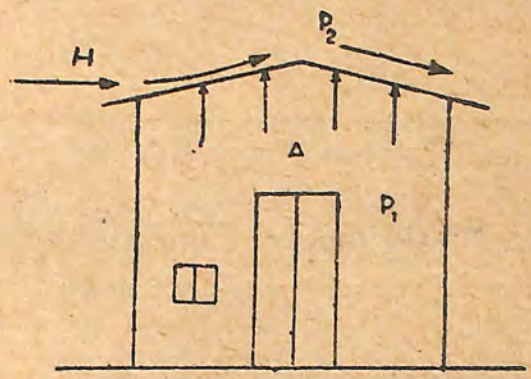


Fig. 12.16 Atmospheric pressure p_1 and pressure p_2 on the top of the roof causes the roof to be lifted, H-wind with high velocity

(ii) In a common form of atomizer, squeezing the bulb spray out oil or scent from the nozzle (Fig. 12.17). The rise of liquid is not due to any vacuum created in the central tube. When the bulb is squeezed the air blows through the central tube with high speed causing a low pressure p_2 inside. The atmosphere pressure p_1 on the liquid

surface of the container pushes the liquid up the steam which then is blown out with the air steam through the nozzle.

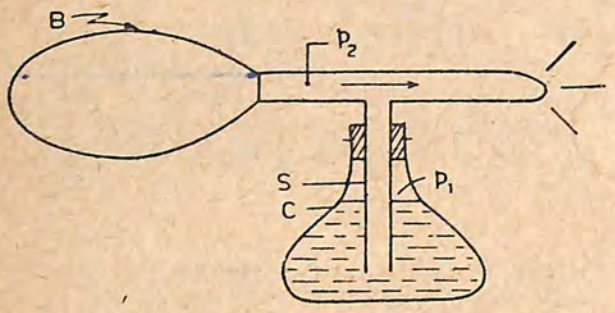


Fig. 12.17 Atomizer; . B-bulb; p_1-atmospheric pressure: C-container; S-stem; P_1-pressure inside

(iii) Lift on an aircraft wing: Figure 12.18 shows the cross-section of the wing of

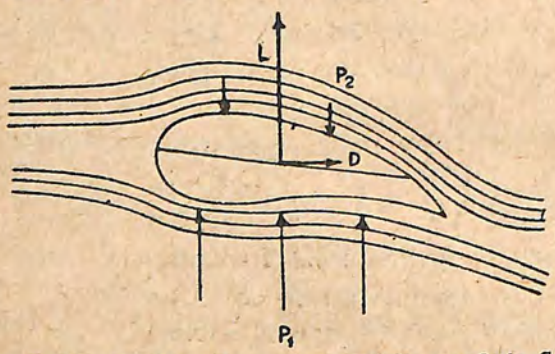


Fig. 12.18 Profile of a wing and lines of air flow

an aircraft. When the aircraft moves, the air flows both from under and above the wing. Aeroplane wings are so designed that the total distance travelled by air flowing over the wing is longer than that of the air flowing under it. Thus, the velocity of air flow above the wing must be higher than the velocity of air flow under the wing. From Bernoulli's theorem we see that the pressure p_2, above the wing is lower than the corresponding pressure p_1, under the wing. The unbalanced pressure causes a force F to act on it. This force can be resolved into two components. The vertical component is responsible for the lift L and the horizontal component provides the drag D. The force acting on it depends on angle of attack. If angle of attack is too great, the streamline flow above and behind the wing breaks down and a complicated system of whirls and eddies known as turbulence is set up. Bernoulli's equation no longer applies. The pressure above the wing rises and the lift on the wing decreases. The plane stalls.

There are various other examples which can be observed by the reader in his daily life where Bernoulli's principle is used.

(iv) Curved path of a spinning ball: When a ball is thrown or hit with a horizontal velocity v and given a spin the path of the projectile of the ball curves more than in a usual spin free projectile. Let a cricket ball moving to the right be given a spin as shown in Figure 12.19 (a). This situation

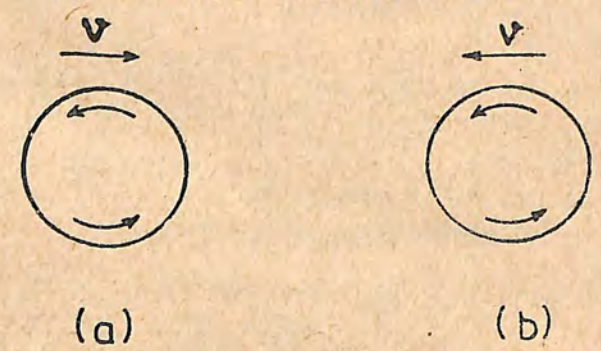


Fig. 12.19 Motion of a spinning cricket ball

is effectively the same as the ball spinning with air wind moving in opposite direction of v (Fig. 12.19 (b). Thus just above the ball the wind has a velocity greater than v as the velocity componant due to spin adds up to v and just below the opposite is the case. A pressure difference of air created

at the top and bottom of the spinning ball, produces an additional unbalanced force. This force deviates the spinning ball from the path of a free projectile. If the spin is in opposite sense pressure difference above and below the ball is reversed and the ball curves more downwards.

Exercises

12.1 A liquid drop of diameter D breaks up into 27 tiny drops. Find the resulting change in energy. $(2\pi D^2\sigma)$

12.2 What would be the gauge pressure inside an air bubble of 0.20 mm radius situated just below the surface of water ? Surface tension of water is 0·07 Nm^{-1}. $(7 \times 10^{-2}\ Nm^{-2})$

12.3 A U tube is such that the diameter of one limb is 0.4 mm and (Hint : unlike soap bubble there is only one surface here) that of the other is 0.8 mm. Find the difference in the level of water in the limbs. $\sigma = 0.07 Nm^{-2}$. (3.4 mm)

12.4 Why is it that a needle may float on clear water but will sink when some detergent is added to water ?

12.5 At what speed will the velocity head of a stream of water be equal to 40 cm of Hg ? (2.8 m/s)

12.6 If a small ping pong ball is placed in a vertical jet of air or water, it will rise to a given height above the nozzle and stay at that level. Explain.

12.7 A pilot-tube is mounted on an aeroplane wing to measure the speed of the plane (Fig. 12.20). The tube contains alcohol and shows a level difference of 40 cm. What is the speed of the plane relative to air (sp.gr. or alcohol = 0.8 and density of air = 1 kgm^{-3}) ? $(56\sqrt{2}$ m/s)

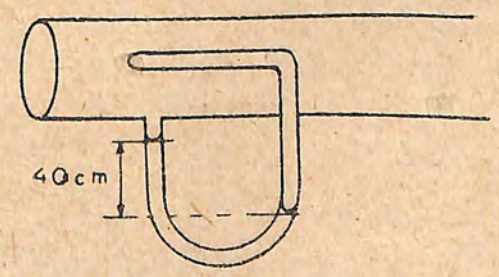


Fig. 12.20

12.8 A glass tube of 1 mm bore is dipped vertically into a container of mercury, with its lower end 2 cm below the mercury surface. What must be the

gauge pressure of air in the tube to blow a hemispherical bubble at its lower end. ($2805.6\ Nm^{-2}$)

12.9 Water stands at a depth H in a tank whose side walls are vertical (Fig. 12.21). A hole is made on one of the walls at a depth h, below the water surface.

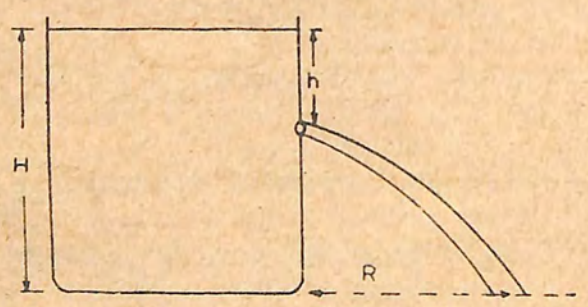


Fig. 12.21

(a) At what distance R from the foot of the wall does the emerging stream of water strike the floor ?

(b) For what value of h this range is maximum ? ($2\sqrt{h(H-h)}$; $h = H/2$)

12.10 With what terminal velocity will an air bubble 0.8 mm in diameter rise in a liquid of viscosity $0.15 Nm^{-2}s$ and specific gravity 0.9 ? What is the terminal velocity of the same bubble in water ? ($209 \cdot 10^{-5}$ m/s)

12.11 A sphere is dropped under gravity through a fluid of viscosity η. Taking the average acceleration as half of the initial acceleration, show that the time to attain the terminal velocity is independent of the fluid density. $\left(V = \frac{2}{9} \frac{r^2 \rho}{\eta} \right)$

UNIT 13

Electricity

13.1 Electric Current

The term current implies some sort of motion. A motion of electric charge constitutes an electric current. This motion may be that of the charge carriers in a medium, like electrons in a metal and ions in a liquid or a gas, or it may be the motion of electrons in a vacuum tube.

Quantitatively, electric current is defined as the rate of flow of charge, given by

$$I = q/t \qquad \ldots (13.1)$$

where I is the current and q is the charge that has passed through a given area in time t.

If the rate of flow of charge does not change with time the current is said to be steady. In many situations, however, the current may be varying with time. Thus in Fig. 13.1, curve (a) represents a steady current while curves (b) and (c) represent varying currents.

By convention, the direction of motion of positive charge is taken as the direction of current. A negative charge moving in one direction is equivalent to an equal positive charge moving in the opposite direction (Fig. 13.2).

In S.I. units the current is measured in amperes, 1 ampere = 1 coulomb/second. It is interesting to note that in practice the electric currents are not measured directly by measuring charge and time but by the effects produced by them, for instance, the magnetic effects.

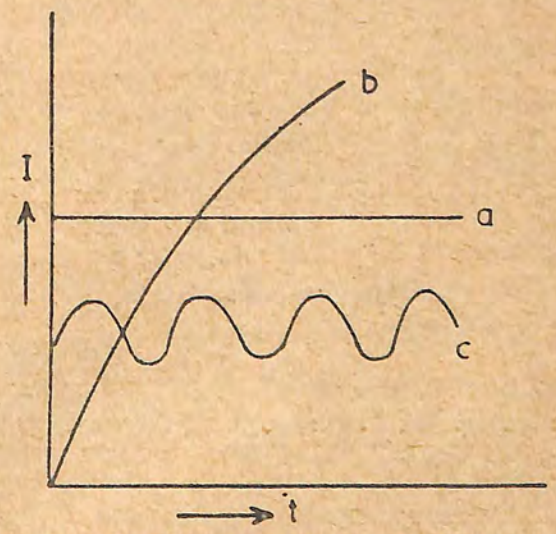


Fig. 13.1 Examples of steady and varying currents

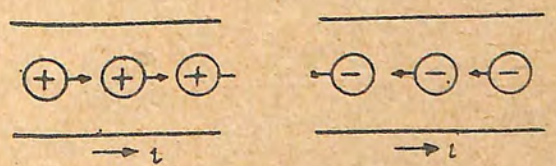


Fig. 13.2 Flow of positive charge is equivalent to flow of negative charge in the opposite directions.

Example 13.1

One billion electrons pass from a point A towards another point B in 10^{-3} seconds. What is the current in ampere ? What is its direction ?

Solution

Charge on 1 electron $= 1.6\times10^{-19}$C

$q = 1.6\times10^{-19}\times10^{9}$

$=1.6\times10^{-10}$C

$t = 10^{-3}$ s.

$I = q/t = 1.6\times10^{-10}/10^{-3}$

$= 1.6\times10^{-7}$Cs^{-1}

$= 1.6\times10^{-7}$ A.

Since the electrons are negatively charged, the direction of current flow is from B to A.

Example 13.2

The electron in the hydrogen atom circles around the proton with a speed of 2.18×10^{6} m/s in an orbit of radius 5.3×10^{-11} m. Find the equivalent current.

Solution

Radius of orbit, $r = 5.3\times10^{-11}$ m.

Circumference of the orbit $=2\pi r$

$=2\pi\times5.3\times10^{-11}$ m

Speed of electron, $v=2.18\times10^{6}$ ms^{-1}

In one second the electron will go around the proton n times, where

$$n = \frac{v}{2\pi r} = \frac{2.18\times10^{6}}{2\times3.14\times5.3\times10^{-11}}$$

The electric current is defined as the quantity of charge that flows past a fixed point in one second. Although it is the same electron going round and round the proton, the quantity of charge that passes across a fixed point on the orbit in one second is given by ne, where e is the electronic charge.

Hence the equivalent current

$$I=ne=\frac{2.18\times10^{6}}{2\times3.14\times5.3\times10^{-11}}\times1.6\times10^{-19}\text{Cs}^{-1}$$

$=1.05\times10^{-3}$ A.

Although we are more familiar with currents in a wire, currents can also flow through liquids and gases. The only requirement is the availability of charge carriers and a difference of potential. A mobile positive charge carrier would then move from a region of high potential to a region of low potential under the influence of the electric field, a negative charge carrier in the opposite direction. In solid conductors, the charge carriers are mostly electrons, in liquids and gases it is the positive and negative ions (in gases, electrons also) that contribute to the current flow.

Consider a pair of charged plates (Fig. 13.3),

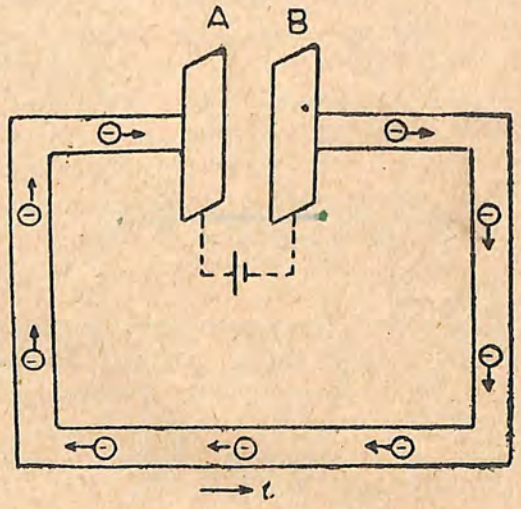


Fig. 13.3 Steady current requires a source of constant potential difference

plate A being at the higher potential. If we connect the plates by a wire, electrons will travel from plate B to plate A and an electric current would be said to flow from A to B, through the wire. Now, the arrival of electrons at the plate A will tend to equalise the potential between the plates and, unless the potential difference is maintained by external means, the current would soon cease to flow. Thus, for steady currents to flow between two points

the potential difference between them must be maintained by a battery or any other device.

Current I through a conductor depends upon the potential difference V applied across it. By changing V, one can study the variation of I with V. Fig. 13.4 show such plots for

Fig. 13.4 V-I plots for (a) metallic conductor (b) vaccum tube (c) electrolyte

(a) a metallic conductor, (b) a vacuum tube, and (c) an electrolyte. The ratio

$$\frac{V}{I} = R \qquad \ldots (13.2)$$

is known as the resistance of the conductor.

We see that for the metallic conductor, the V-I curve is a straight line i.e. the resistence R of the conductor is independent of V. In this case we say that the conductor obeys Ohm's law. All such conductors which show this kind of linear behaviour are known as ohmic resistors. Others, for which the V-I curve is not a straight line are called non-ohmic resistors. Both kinds are used in electric circuits for various purposes.

13.2 Current Flow in a Metallic Conductor : Drift Velocity

Let us examine in some detail the mechanism of current flow in a metallic conductor.

In a metal in solid state, atoms occupy places in a fixed arrangement. In this state some electrons are free, leaving the atoms as positively charged ions. If we assume that there is one free electron per atom, there will be of the order of 10^{28} free electrons per cubic metre. Since they possess thermal energy, they are in a state of continual motion, much like the molecules of a gas. They are colliding with atoms, gaining and losing kinetic energy. At room temperature they move with velocities of the order of 10^5 ms^{-1}, but these velocities are randomly distributed in all directions. Therefore, the average velocity of electrons, given by

$$u = \frac{u_1 + u_2 + u_3 + \ldots + u_N}{N} = 0$$

where u_1, u_2 etc. are velocities of individual electrons and N is the total number of electrons.

In other words, there is no net flow of charge in any direction. It may be noted that average speed is not zero.

When a potential difference is applied across the conductor, an electric field E is established inside it. Due to the electric field the electrons experience coulomb force in a direction opposite to the field and they undergo an acceleration ($a = -\frac{eE}{m}$). The forces are experienced by positive ions also but they are unable to move as they are heavy and tightly bound in the metal. Between two successive collisions an electron thus acquires, in addition to its thermal velocity, a velocity component in a direction opposite to E. This gain in velocity due to the field E, however, is very small and is lost in the next collision with an atom.

If we looked at any arbitrary instant, we would find an electron moving with a velocity $u_1 + at_1$, where u_1 is its velocity due to thermal motion and at_1 is the velocity acquired under the influence of the external field E, t_1 being the time that has elapsed since the last

→ Ohm's law

collision. The average velocity of all electrons will be given by

$$v = \frac{(u_1 + at_1) + (u_2 + at_2 + .) \ldots\ldots + (u_N + at_N)}{N}$$

$$= \frac{u_1 + u_2 + \ldots\ldots u_N}{N} + \frac{a\,(t_1 + t_2 + \ldots\ldots t_N)}{N}$$

$$= 0 + at, \text{ where } t = \frac{t_1 + t_2 + \ldots\ldots t_N}{N}$$

t represents the average time elapsed since each electron suffered its last collision, it is called the relaxation time and is $= 10^{-14}$ s. It is a characteristic of the given conductors. Denoting v by v_d

$$v_d = -e\,E\,t/m \qquad \ldots (13.3)$$

v_d is called the drift velocity of electrons.

We may say that before the application of the field, the average velocity of the electrons is zero and there is no net transport of charge in any direction. After the field is applied, electrons acquire a constant average drift velocity in a direction opposite to E and there is a net transport of negative charge in that direction which is equivalent to an electric current flowing in the direction of E. A typical value of v_d is 1 mm/s, which is very small compared to the random velocity (10^5 m/s.) But, since the number of electrons is very large, even this small drift of electrons adds up to a sizeable current.

One sometimes tends to confuse between the drift speed of electrons and the speed of propagation of electrical effects. It may look surprising that although the drift velocity of electrons is so low, an electric bulb turns on almost immediately when it is switched on. The fact is that the propagation of an electric impulse takes place with a speed which is of the order of speed of light ($=3\times10^8$m). It is somewhat like applying a pressure to one end of a long water filled tube. As soon as the pressure is applied a pressure wave is transmitted rapidly along the tube and when it reaches the other end the flow of water starts. The water inside the tube also starts moving forward but this speed is much slower than the speed of pressure wave. Similarly in a circuit, free electrons are present everywhere. When a potential difference is applied to the circuit, an electric field gets established throughout the circuit almost with the speed of light. As soon as the electric field is established the electrons begin to drift under its influence and a current flows in the circuit.

Resistance

Consider a conductor of length l and of uniform cross section A, across which is applied a potential difference V, Fig. 13.5. The electrons start drifting towards left with a

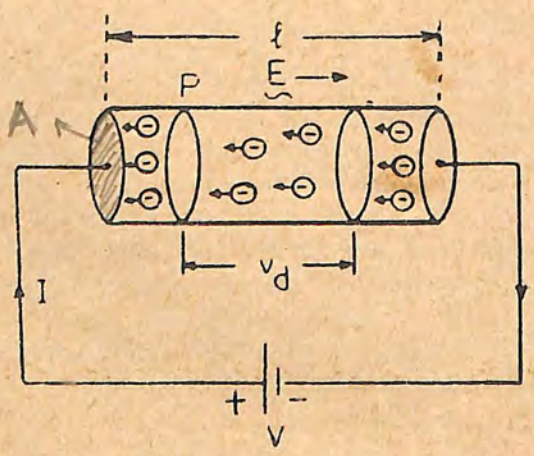


Fig. 13.5 Flow of current through a conductor

velocity v_d. All those electrons which lie within a distance v_d on the right of a cross section at P, would have passed on to the other side in one second. The net transport of charge through the section at P in 1 second or the current I is, therefore, given by

$$I = enAv_d \qquad \ldots (13.4)$$

where n is the number of free electrons per unit volume in the conductor. It is easily seen;

using equation (13.3), that

$$I = enA.\ \frac{eEt}{m}$$

$$= \frac{(e^2nAt)}{ml} V, \text{ since } E = V/l.$$

Hence, the resistence,

$$R = \frac{V}{I} = \frac{ml}{e^2nAt} \qquad (\ldots 13.5)$$

R is a constant for the given conductor . We can also write

$$R = \rho\, l/A, \qquad \ldots (13.6)$$

Where $\rho = m/e^2nt$ ... (13.7)

is called the resistivity of material of the conductor. The resistivity is a constant of the material of the conductor. It can be verified that units of ρ are ohm-metre.

EXAMPLE 13.3

What is the drift velocity of electrons in a copper conductor having a cross sectional area of $5\times10^{-6}m^2$ if the current is 10A. Assume there are 8.0×10^{28} electrons/m^3.

Solution

$$I = en\,AV_d$$

Here, $e = 1.6\times10^{-19}C$, $n = 8.0\times10^{28}/m^3$,

$A = 5\times10^{-6}m^2$, $I = 10$ A

$$V_d = \frac{I}{enA}$$

$$= \frac{10}{1.6\times10^{-19}\times8.0\times10^{28}\times5\times10^{-6}}$$

$$= 1.56\times10^{-4} \text{ m/s.}$$

13.3 Heating Effects of Current : Joule's Law

It is a matter of common observation that when a current flows in a conductor heat is developed. This can be understood on the basis of the microscopic picture presented above. We have seen that the motion of electrons in a conductor under the influence of the field is hampered by their frequent collisions with the atoms; this being the basic cause of resistance. In between the collisions, the electrons moving through the field, acquire kinetic energy in addition to their own kinetic energy of thermal motion. In the collisions this energy is shared with the atoms of the metal with a consequent increase in the average kinetic energy of the atoms. The conductor is thus heated by the flow of electric current through it.

Consider the portion AB of a circuit (Fig. 13.6) through which a current I is flowing in the

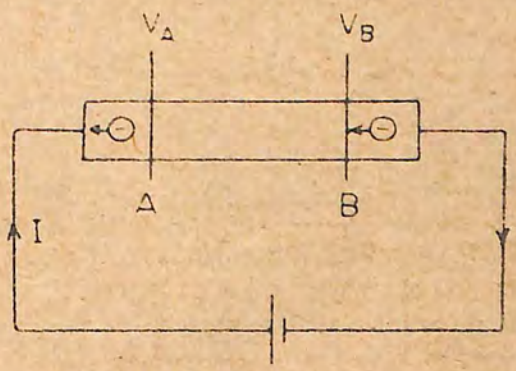


Fig 13.6 For calculation of power dissipation in a conductor

direction A to B. Electrons enter at B at potential V_B and leave at A at potential V_A. Since electrons are negatively charged and $V_A > V_B$ the electrons entering at B possess more potential energy than those leaving at A, while their kinetic energies are the same. The electrons thus lose potential energy while passing from B to A. It is this difference in potential energy which is transferred to the conductor as heat energy. An expression for the same can be drived as follows:

The potential energy of an electron at B $= (-e)\, V_B$.

The potential energy of an electron at A $= (-e)\, V_A$.

Number of electrons which enter the portion AB at B and leave it at A, in time $dt = I dt/e$

The loss in potential energy of electrons in time dt is

$$dw = \frac{Idt}{e}\ [(-e)V_B - (-e)V_A]$$
$$= Idt\,(V_A - V_B)$$
$$= Idt\,V_{AB} \text{ where } V_{AB} = V_A - V_B$$

This energy appears as heat. Therefore, the rate at which the electric energy is converted into heat, that is, power P is given by

$$P = \frac{dw}{dt} = V_{AB}\,I \qquad \ldots (13.8)$$

Since $V_{AB} = IR$, where R is the resistence of the portion AB of the conductor, we have

$$P = V_{AB}I = (IR)\,I = I^2R \qquad \ldots (13.9)$$

This is Joule's law of heating. If V_{AB} is in volts and I in amperes, then P is in watts.

$$1 \text{ amp} \times 1 \text{ volt} = \frac{1 \text{coulomb}}{1 \text{second}} \times \frac{1 \text{ joule}}{1 \text{coulomb}}$$
$$= \frac{1 \text{ joule}}{1 \text{second}} = 1 \text{ watt.}$$

Heat developed in time t is given by

$$Q = Pt = I^2 Rt \text{ joules} = \frac{I^2Rt}{4.2} \text{calories} \qquad \ldots (13.10)$$

Since 1 calorie = 4.2 joules.

The kilowatt hour : If P is power in watts and t is time in seconds, the energy consumed is

$$W = Pt \text{ joules}$$

For consumption of electric energy, joule is considered to be too small a unit for practical purposes such as calculating monthly electricity bill. The practical unit for electric energy is *kilowatt-hour*, defined by

$$W = Pt$$

where, P is in kilowatts, t is in hours and W in kilowatt hours. Since 1kw = 1000w, and 1 hour = 3600 seconds, we have,

$$1\text{kwhr} = 1000 \times 3600 = 3.6 \times 10^6 \text{ joules.}$$

EXAMPLE 13.4

An electric heater is rated at 800 watts.

(i) If the voltage is 200 volts what is the value of current ?

(ii) How much time would be required to heat 1 litre of water from 20°C to 100°C ?

Solution

(i) $I = P/V = 800/200 = 4.0\text{A}$

(ii) Heat required = mass of water × specific heat of water × rise in temperature
= (1000gm) (1 cal/gm°C) (80°C)
= 80,000 cal. = 80,000 × 4.2 joules.

Therefore, $t = \frac{Q}{P} = \frac{80{,}000 \times 4.2}{800} = 420\text{s.}$

EXAMPLE 13.5

What voltage drop will be there across a 1kw electric heater element whose resistance when hot is 40 ohms ?

Solution

$$P = I^2R = (V^2/R^2)R = V^2/R$$

$$\therefore V = (P \times R)^{\frac{1}{2}} = (1000 \times 40)^{\frac{1}{2}} = 200 \text{ volts.}$$

13.4 Variation of Resistance with Temperature

It has been stated before that the atoms in the metal occupy fixed positions. They are, however, not stationary but vibrate to and fro about their fixed positions. When a metal gets heated there is an increase in the average kinetic energy of the atoms and they vibrate more vigorously. This leads to more collisions of electrons with atoms and a consequent decrease in the value of t. Also at highert emperature the thermal speed of electrons is more and this further reduces the time between collisions. A lower value of t, therefore, results in an increase in the resistivity of the conductor equation (13.7.)

Experimentally it is found that to a first approximation

$$R_t = R_0 (1 + \alpha t) \quad \ldots (13.11)$$

where R_t and R_0 are resistances at temperatures t°C and 0°C respectively, and α is known as the temperature coefficient of resistance For metals α is a positive coefficient and for copper its value is 0.0040 per degree. For some alloys, like constantan and manganin the value of α is very small (0.00001) and these are used in making standard resistances. Carbon and other semiconductors display a negative coefficient i.e. their resistance decreases with temperature. α is negative for electrolytes also. It is beyond the scope of this book to go for an explanation of negative α.

13.5 Thermo-electric Effects

Fig. 13.7 shows two pieces of wire, one copper and the other iron, joined together at

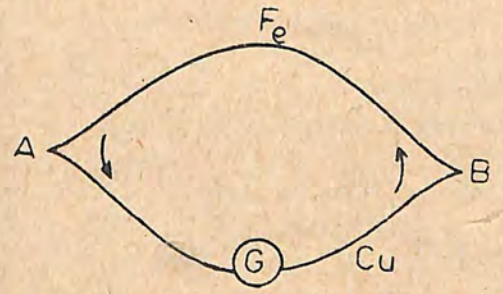


Fig. 13.7 A thermo couple : A is cold Junction and B is hot junction

their ends A and B. A low resistance galvanometer G is included in the circuit. When one junction is heated and the other kept at a constant temperature, a small electric current flows around in the circuit in the direction of arrows. This is known as Seebeck effect after its discoverer. The effect is exhibited by any two dissimilar metals joined in this manner. Seeback investigated the effect for a number of metals and arranged them in a series such that when any two of them form a circuit the current flows at the hot junction from the metal occurring earlier in the series to the one occurring later. Some of the substance in the series are Bi, Ni, Pt, Cu, Rh, Ir, Fe, Sb.

Current produced in this way is known as thermoelectric current and the pair of metals forming the circuit is called a thermocouple. Evidently the thermoelectric current flows because an e.m.f. (electromotive force) has been set up in the circuit purely due to thermal effect. This e.m.f. is known as thermo e.m.f. and can be measured by using an instrument called a potentimeter. It is a normal practice to measure this e.m.f. by keeping one junction at a fixed temperature, very often at O°C, and varying the temperature of other junction. A typical curve is shown in Fig. 13.8.

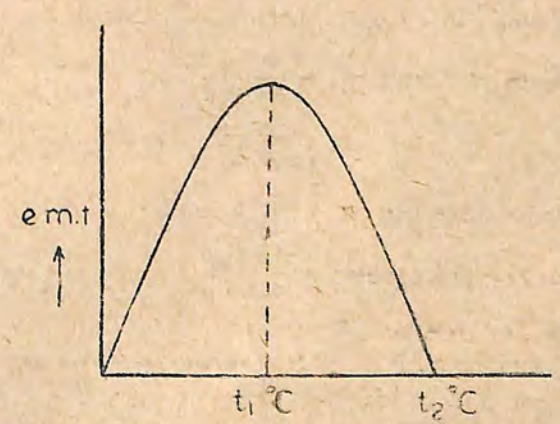


Fig. 13.8 Variation of thermo e.m.f. with temperature

The maximum e.m.f. is obtained at t_1°C which is called the neutral temperature of thermocouple. At t_2°C the thermo e.m.f. falls to zero value and reverses its direction. This temperature called the inversion temperature, is not unique but depends upon the temperature of cold junction. It is always as much above the neutral temperature as the cold

junction is below it. For Cu-Fe couple, the neutral temperature is about 300°C.

13.6 Temperature Measurement by Thermocouples

Thermocouples are widely used for temperature measurement. If the temperature of cold junction is fixed, the thermal e.m.f. of a thermocouple is a function of the temperature of hot junction alone. For a considerable range of temperature this e.m.f. is a linear function of temperature. This fact can be used for measuring the temperature at the site of the hot junction if the couple is previously calibrated. Frequently the galvanometer in the circuit is calibrated to read directly the temperature. The choice of thermocouple wires depends upon the range of temperature required. The copper-constantan couple, consisting of copper and an alloy called constantan, is used to measure temperature ranging from −190°C to 300°C. Thermocouples of platinum and a platinum-rhodium alloy are used for measuring temperatures upto 1600°C.

As a temperature measuring device, the thermocouple has many merits. It is accurate. The copper-constantan couple, for example, gives an e.m.f. of 40 microvolts per degree temperature difference between its junctions. This e.m.f. can be measured with an accuracy of about 1 microvolt. The temperature differences, therefore, can be measured with an accuracy of about 1/40 degree. Another great advantage of a thermocouple is the smallness of its test junction. This makes it particularly suitable for measuring temperature in very small regions and cavities. Because of its small mass the test junction comes quickly into thermal equilibrium with the surroundings. Use is made of this fact in measuring variations in the temperature in various parts of the body of animals and insects.

Thermopile

The e.m.f. generated by a thermocouple is small, of the order of few millivolts. But, if a number of thermocouples are connected in series the e.m.f.s are added up. Such a combination is called a thermopile and is used to measure the intensity of heat radiation.

One set of junctions (A) is exposed to the radiation, the other set (B) is protected by an insulating lid (Fig. 13.9). The exposed ends

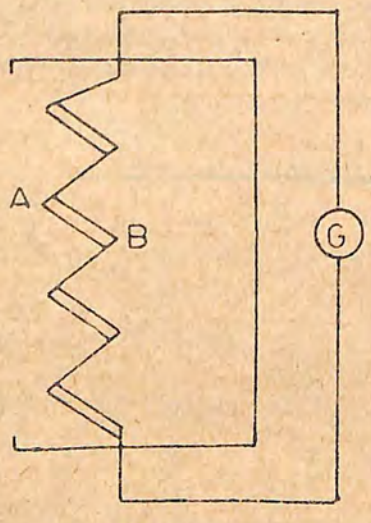


Fig. 13.9 Thermopile

are blackened to increase the absorption of radiation falling on them. The thermocouples are made of antimony and bismuth strips. The galvanometer deflection is proportional to the intensity of falling radiation.

13.7 Magnetic Effects of Electric Current

Before we discuss the magnetic effects of current as such, we discuss the nature and meaning of electric and magnetic fields.

The forces and the fields

When a body exerts a force on another

body and vice versa, the two bodies are said to interact with each other. Thus, the earth and the moon interact to exert mutal forces on each other. These forces are given by Newton's law of gravitation. Another kind of interaction called electric interaction occurs between two charged bodies, the force that one charged body exerts upon the other is given by Coulomb's law. We are also aware of the force that one magnet exerts on another. This is called magnetic interaction. However, this is not a new kind of interaction. As we shall presently see, the magnetic forces can be explained in terms of electrical forces between moving charges. The common interaction which describes both electric and magnetic forces is called, aptly the electromagnetic interaction.

In the example cited above, there is one common feature—these are examples of 'action at a distance.' The force is transmitted from one object to another without any direct contact. In contrast most of the forces that we come across in daily life seem to involve direct physical contact e.g. pushing a body, pulling a string, etc. How do the gravitational, the magnetic and the electric forces transmit themselves from one object to another ?

The interaction between two bodies through intervening space can be understood by introducing the concept of 'field'. We illustrate it by taking the examples of force between two charges.

The experimental fact that an electric charge experiences a force due to another charge located some distance away can be understood as follows. We say that one of the charges sets up an electric field in the surrounding space, this field acts on the other charge and produces the observed force. Since the force is mutual, we may regard any one of the charges to be a source of the field which acts upon the other. The electric field thus acts an intermediary for interaction between two charged bodies.

Similarly, the two masses are supposed to interact via a gravitational field and the two magnets through a magnetic field. The origin of magnetic field can, however, be traced back to the moving charges as discussed later.

The electric field

Consider a charge q situated at some point in space. It creates an electric field around it. We call q, the source charge. If another charge q_o, which we call the test charge, is placed at some point in this field (Fig. 13. 10(a), it experiences a force given by

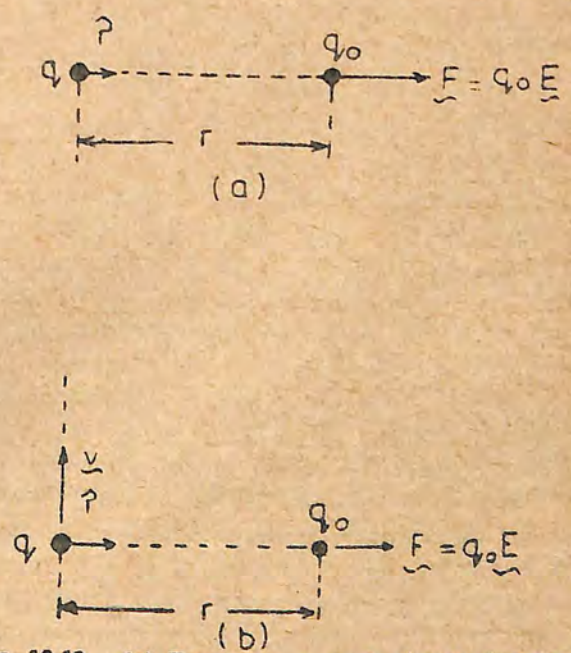


Fig. 13.10 (a) Force on a test charge due to a stationary charge
(b) Force on a test charge due to a moving charge ($V << c$)

$$\mathbf{F} = q_o \mathbf{E} \qquad \ldots 13.12)$$

where E is the electric field at that point.
To relate the value of E to the source charge

we use Coulomb's law which gives the force F between q and q_o as

$$F = \frac{q\, q_o}{4\pi\epsilon_o\, r^2}\hat{r} \qquad \ldots (13.13)$$

where r is the unit vector from q to q_o and ϵ_o ($=8.85\times10^{-12}C^2/Nm^2$) is the permitivity of free space. The electric field E, due to q at the site of q_o is, therefore, given by

$$E = \frac{q}{4\pi\epsilon_o\, r^2}\hat{r} \qquad \ldots (13.14)$$

In the above discussion the source of the field was a stationary charge. It is found that a moving charge is also a source of electric field. If the charge q is moving (Fig 13.10 b) then the force on a stationary charge q_o is still given by $F=q_oE$, where E is the electric field due to q at the site of q_o at the instant of observation. For small velocities* of q, the field is the same as due to a stationary q, given by equation (13.14).

The electric field obeys superposition principle. The field produced at a certain point in space when there are number of sources is the vector sum of the fields that each source would individually produce if it alone were present.

The force equation, $F=q_o E$, enables us to define the electric field at a point without reference to its sources. If a test charge placed at rest at a point in space experiences a force, we say an electric field exists at that point. The strength of the field is the force per unit charge on the test charge and its direction is the same as that of the force acting on a (positive) test charge, In symbols,

$$E = \frac{F}{q_o} \qquad \ldots (13.15)$$

An important property of the electric field is this: The force that it exerts on a test charge is independent of the velocity of test charge. Thus, if there is an electric field E at a point, the force exerted on a test charge q_o at that point will be $q_o E$, irrespective of the fact whether the charge q_o is at rest or in motion.

The Definition of Magnetic Field

Laboratory experiments show that two parallel wires carrying currents in the same direction attract each other (Fig. 13.11(a), if

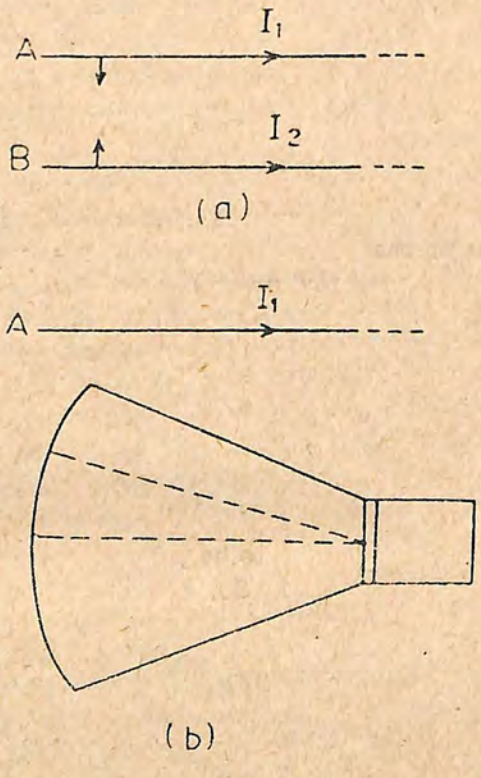


Fig. 13.11 (a) Magnetic interaction between parallel wire carrying currents
(b) Deflection of an electron beam by a magnetic field

* For high velocities (of the order of velocity of light) the field becomes somewhat complicated but we need not to go into these details here.

the currents are in opposite directions, they repel. It is convenient to imagine, that one current produces some kind of field which exerts a force on the other current. Let us investigate the nature of this field.

This field cannot be electrostatic as the wires have no net charge on them because they have equal number of electrons and positive ions. Moreover, if we remove wire B and place a charged conductor there, it experiences no force. This field, therefore, does not interact with stationary charges.

Further, if we switch off i_2, the mutual force of attraction (or repulsion) is also switched off, although current i_1 and its field are still there. It follows, therefore, that the field due to i_1 acts on the wire B only when there is a current through it, that is when there is a flow of charge through wire B.

To test whether the field acts only upon the charges moving in a conductor or it also acts on charges moving freely we replace B with an evacuated discharge tube (Fig. 13.11(b) and find that the electron beam gets deflected (i.e. force acts on the moving charges).

The magnitude of deflection varies with the direction of motion of the charges in the beam. It is found that there is a direction for which the deflection is zero and at right angles to it, the deflection is a maximum. Also, the deflection is found to be proportional to the speed of electrons. So, here we have a field which exerts, on a moving charge, a force, which is velocity dependent. We call it the magnetic field and denote it by the symbol B.

If we assume that the direction of magnetic field coincides with the direction of motion of charges when the deflection is zero, we find that, in general, the force on a charge q moving with velocity v can be described by the experssion.

$$F = qvB \sin\phi \qquad \ldots (13.16)$$

where ϕ is the angle between B and v. Also, it is found that force F is in a direction perpendicular to both B and v. This enables us to express F, in vector notation as

$$\mathbf{F} = q(\mathbf{v} \times \mathbf{B}) \qquad \ldots (13.17)$$

The relationship between F, v and B is shown in Fig. 13.12. It is governed by the

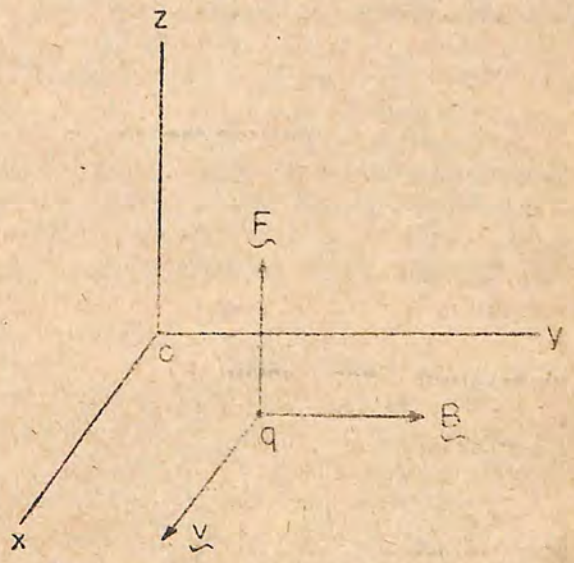


Fig. 13.12 Shows the direction of force on a charge moving in a magnetic field

usual rule for a vector product the direction of F will be the one in which a right handed screw will advance by a rotation of v towards B through the smaller angle between them. Then, if v and B lie in the x—y plane as shown, F will point out in the z direction. [Note that for a beam of electrons q is negative and it follows from equation (13.17) that F has the opposite direction to that of $(\mathbf{v} \times \mathbf{B})$.]

Unit of B : In S. I. system, unit of magnetic field B is called Tesla. In the equation,

$$B = \frac{F}{qv\sin\phi} \qquad \ldots (13.18)$$

if $F = 1$ newton
$Q = 1$ coulomb
$v = 1$ metre/second
$\phi = 90°$

Then, $B = 1$ newton-second/coulomb-metre = 1 tesla, that is, a charge of 1 coulomb moving with a speed of 1m/s at right angles to a magnetic field of strength 1 tesla will experience a force of 1 newton.

A field of 1 tesla is a very strong magnetic field. The field in the neighbourhood of common permanent magnet is 0.1 tesla. The earth's magnetic field on the earth's surface is 5×10^{-5} tesla. Very often the magnetic fields are expressed in terms of a smaller unit, called the gauss. 1 gauss $= 10^{-4}$ tesla.

In some books B is called the magnetic induction.

Source of the magnetic field

We have seen that a current carrying wire produces a magnetic field. If the current is switched off, the magnetic field disappears.

It immediately suggests that the magnetic field is the result of the motion of electrons in the wire. One may go further and say, in general, a moving charge is a source of magnetic field.

When there is no current in the wire, the electrons are in thermal motion. Though each electron acts as the source of a magnetic field, these fields cancel out since electrons are moving in a random manner. There is no net magnetic field as there is no net flow of charge in any direction.

13.8 Biot-Savart Law – Magnetic Field due to Some Current Distributions

A magnetic field is produced by a moving charge. We are often interested not in the field of an isolated moving charge but that of a current in a conductor. Each of the moving charges within the conductor contributes to the magnetic field outside. Experiments show that the direction and magnitude of B due to any current depend upon the current direction and the relative position of the observation point. It is more convenient to think that the field is made up of contribution from various segments of the conductor called the current elements rather than from individual charges, in such a case.

The Fig. 13.13 shows a typical current

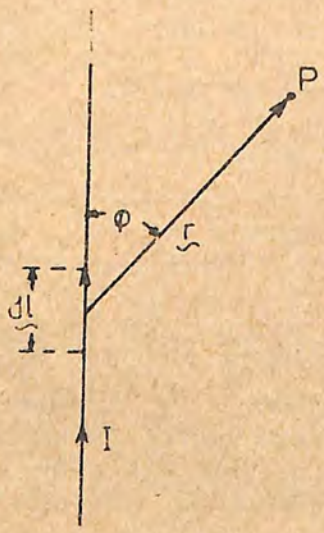


Fig. 13.13 Biot-Savart law

element of length dl of a conductor carrying a current I. The magnetic field dB at the point P, due to the current dl is then given by Biot-Savart Law* :

$$dB = \frac{\mu_0}{4\pi} I \frac{dl \times r}{r^3} \quad \ldots (13.19)$$

* This law was arrived at empirically, by Biot and Savart, in 1820, to provide a consistent explanation of numerous experiments on the magnetic effects of current.

where μ_0 is a constant called the permeability of the vacuum, its value being $\mu_0 = 4\pi \times 10^{-7}$ webers/ampere metre.

The magnitude of the field is given by

$$dB = \mu_0 \frac{Idl\sin\phi}{4\pi r^2} \quad \ldots (13.20)$$

where ϕ is the angle between the vector dl and r, r being the position vector of P relative to the element.

The direction of dB is that of vector $dl \times r$. In the figure 13.13, dB at the point P is directed into the page at right angle to the plane of the figure. It is easy to make use of Biot-Savart law to get the field in case of regular geometry of current distributions.

Field at the centre of a circular coil

Consider a loop of wire of radius r carrying a current I, as shown in Fig. 13.14. The loop

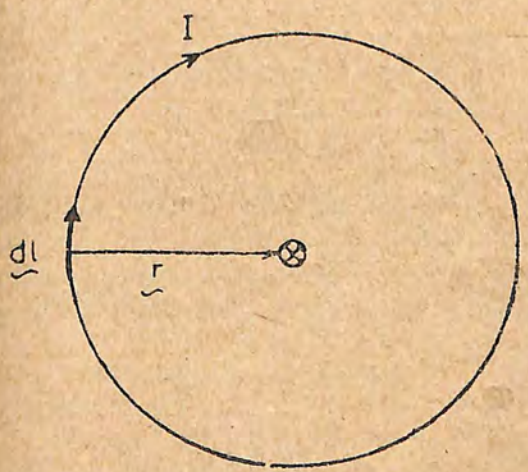


Fig. 13.14 Magnetic field at the centre of a circular current

lies in the plane of paper. We wish to calculate the magnetic field at the centre of the loop.

Consider the element dl of the loop. The magnetic field dB at the centre O due to this current element is given by Biot-Savart law,

$$dB = \frac{\mu_0}{4\pi} \frac{I.\ dl \times r}{r^3}$$

The direction of the field is perpendicular to the plane of the loop and points into the plane of paper. It is shown by the sign $\otimes$. Since the angle between dl and r is 90°, the magnitude of dB, using equation (13.20) is given by

$$dB = \frac{\mu_0 Idl}{4\pi r^2} \quad \ldots (13.21)$$

If we divide the loop into a large number of such elements we find that the field at O due to each element points into the plane of paper and its magnitude is given by equation (13.21). The total field at O is given by

$$B = \int dB = \int \frac{\mu_0 Idl}{4\pi r^2} = \frac{\mu_0}{4\pi}\frac{I}{r^2}\int dl$$

Now, $\int dl$ is just the total length of the loop wire given by $2\pi r$. Hence

$$B = \frac{\mu_0 I}{4\pi r^2}\, 2\pi r = \frac{\mu_0 I}{2r}$$

If instead of a single loop, there is a coil of n turns, all wound over one another, then

$$B = \frac{\mu_0 In}{2r} \quad \ldots (13.22)$$

Field on the axis of the circular coil

The field at a point on the axis of the coil at a distance x from its centre can be shown to be

$$B = \frac{\mu_0 nIr^2}{2(r^2+x^2)^{3/2}} \quad \ldots (13.23)$$

The magnetic field lines due to a circular coil carrying current are shown in Fig. 13.15.

Field due to a solenoid

In Fig. 13.16 are shown the magnetic field lines due to a cylinderical coil carrying a current. Such a coil is normally known as a solenoid.

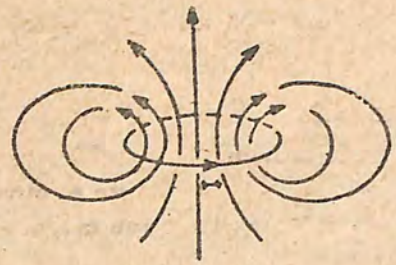


Fig. 13.15 Magnetic field lines due to a circular current

The field is fairly uniform in the interior of the coil and is given by

$$B=\mu_o In \qquad \ldots (13.24)$$

where I is the current and n the number of turns/unit length

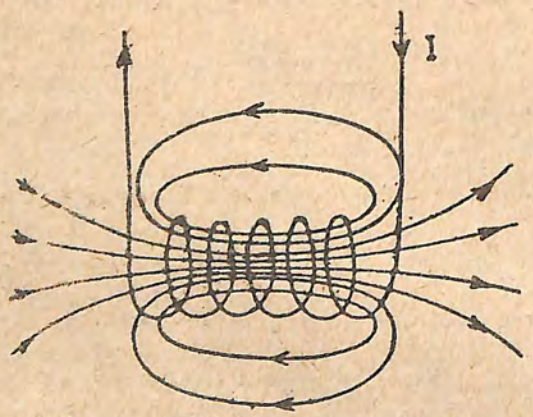


Fig. 13.16 Magnetic field lines due to a solenoid

Field due to a straight current

Fig. 13.17 shows the field lines due to a current I flowing in a straight wire. The field is symmetrical about the current and its value at a point, distant r from the current is given by

$$B=\frac{\mu_o I}{2\pi r} \qquad \ldots (13.25)$$

The field lines are circular and lie in a plane perpendicular to the current.

How to find out field direction ?

The direction of magnetic field in the above

Fig. 13.17 Magnetic field lines due to a straight current

cases can be found out with the help of the following rules :

(i) *For circular current :* Curl the fingers of the right hand in the direction of the current, the stretched thumb then points in the direction of the field.

(ii) *For straight current* Grasp the wire in the right hand with the thumb pointing in the direction of current, the field lines then follow the direction of curled fingers around the current.

EXAMPLE 13.6

Calculate the magnetic field due to a circular coil of 200 turns, radius 0.1m, carrying a current 5A, (i) at a point on the axis of coil distance of point being 0.20m from the centre of coil (ii) at the centre of the coil.

Solution

(i) $B=\dfrac{\mu_o\ nr^2 I}{2\ (r^2+x^2)^{3/2}}$

$\mu_o=4\pi\times10^{-7}$, $n=200$, $r=0.1$m, $x=0.2$m, $I=2$A. Hence

(i) $B=\dfrac{4\pi\times10^{-7}\times200\times0.1\times0.1\times5}{2\times(0\ 1^2+0.2^2)^{3/2}}$

$=5.62\times10^{-4}$ tesla.

(ii) $B=\dfrac{\mu_o nI}{2r}=\dfrac{4\pi\times10^{-7}\times200\times5}{2\times0.1}$

$=62.8\times1^{-4}$ tesla.

13.9 Motion of a Charged Particle in a Uniform Magnetic Field

Fig. 13.18 shows a charged particle of mass m and charge q moving perpendicular to a uniform magnetic field B. It experiences a force $F=q(v\times B)$, where v is the velocity of the particle. Since $F\perp v$, the magnetic force will not speed up or slow down the particle, but will only change the direction of its motion. The particle will move with a constant speed

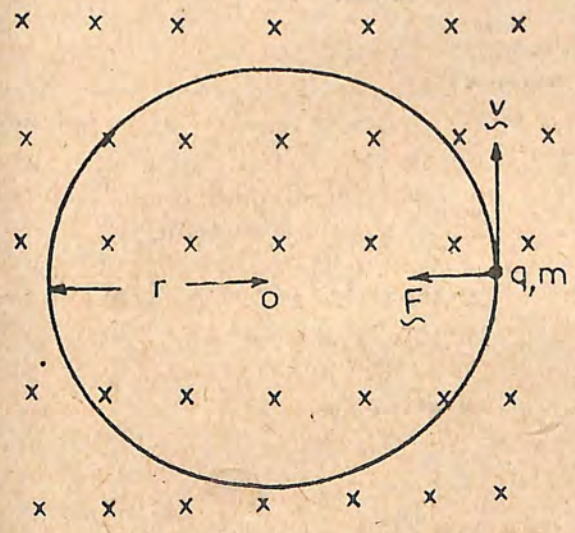


Fig. 13.18 Motion of a charged particle in a uniform magnetic field

but its path will be continually deflected in a way so as to form a circular trajectory.

The force $F=qvB$ gives rise to an acceleration $a=qvB/m$, which is constant in magnitude and perpendicular to velocity. This is the centrepetal acceleration needed to keep the particle in circular motion. Thus

$$\frac{qvB}{m}=\frac{v^2}{r} \qquad \ldots (13.26)$$

where r is the radius of the circle. It may be noted that the circle lies in the plane perpendicular to B.

EXAMPLE 13.7

Electrons moving at right angles to a uniform magnetic field complete a circular orbit in 10^{-9} seconds. What is the magnitude of the magnetic field ?

Solution

Time for completing one orbit is given by

$$t=2\pi r/v$$

Using eq. (13.26), $t=\frac{2\pi r}{v}=\frac{2m\pi}{qB}$

or $B=2\pi\ m/qt$

Now $m=9.1\times10^{-31}$kg., $q=1.6\times10^{-19}$C, $t=10^{-9}$s Hence

$$B=\frac{2\pi\times9.1\times10^{-31}}{1.6\times10^{-19}\times10^{-9}}$$
$$=3.6\times10^{-2}\text{ tesla.}$$

13.10 Force on a Current Carrying Conductor

When a current carrying conductor is placed in a magnetic field, the force experienced by the free electrons moving in the conductor is transmitted to the conductor as a whole. Let a current I flow through a conductor AB placed in a uniform magnetic field B (Fig. 13.19). The magnitude of current is given by

$$I=enAv_d$$

where n is the density of free electrons, v_d their drift velocity and A the area of cross section of the conductor.

Now, each free-electron experiences a force $(-e)(v_d\times B)$. In a length dl, the number of free electrons is nAdl. The total force experienced by the element dl is, therefore, given by

$$dF=nAdl\,(-e)(v_d\times B)$$
$$=-nAe(dlv_d\times B)$$

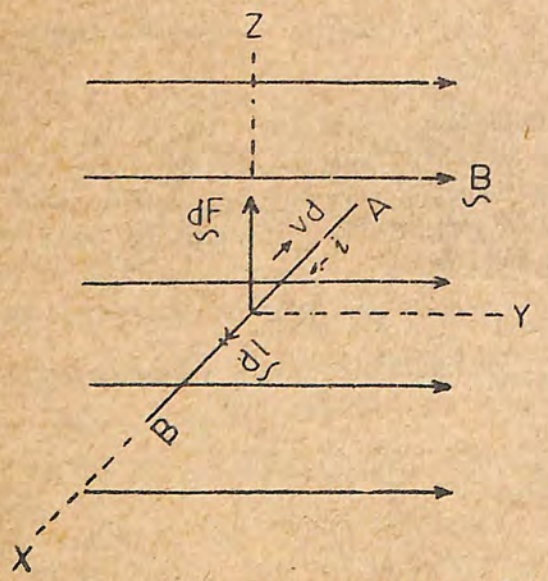


Fig. 13.19 Force on a current carrying conductor

If we represent the element of length by vector dl pointing in the direction of current, then v_d and dl point in opposite directions and the term dlv_d may be replaced by $-v_d dl$. Hence

$$dF = -nAe(-V_d\, dl \times B)$$
$$= nAev_d\, (dl \times B)$$
$$= I\,(dl \times B) \quad \ldots (13.27)$$

If B is uniform over the whole length l of the conductor, the force experienced by the conductor will be

$$F = I\,(l \times B) \quad \ldots (13.28)$$

The direction of the force is given by the vector $l \times B$ where l vector is taken in the direction of the current. If the conductor is held at right angles to the field as in Fig. 13.19, the magnitude of force is given by

$$F = I\,lB \text{ newtons} \quad \ldots (13.29)$$

where I is in amperes, l in metres and B in tesla. The force is perpendicular to both, the conductor and the field. The relationship between the three vectors F, I and B may be described by the so called left-hand rule : Hold the thumb and the first two figures of the left hand mutually at right angles. Point the fore-finger in the direction of the magnetic field, centre finger towards current in the conductor, then the thumb indicates the thurst (force) on the conductor.

Force between parallel wires carrying current

Let AB and CD be two long parallel wires and P a point on AB (Fig. 13.20). The magnetic field at P caused by current I_2 in CD, is

$$B = \frac{\mu_o I_2}{2\pi r} \text{ tesla}$$

Fig. 13.20 Calculating the force between parallel currents

The field B is directed out of the page and is at right angles to AB. The force per metre at P is

$$F = BI = \frac{\mu_o I_1 I_2}{2\pi r} \text{ newton/metre.}$$

The force is at right angle to AB, towards CD, in the plane of paper. The same formula will be obtained for the force exerted on CD. The force will be that of repulsion if the currents are anti-parallel.

Definition of ampere

When $I_1 = I_2 = 1$ ampere and $r = 1$ metre.

$$F = \frac{\mu_o}{2\pi} = 2 \times 10^{-7} \text{ newton.}$$

This is the basis of the definition of ampere. The ampere is the strength of that current which when flowing in two parallel infinitely long conductors of negligible cross section placed in vacuum at a distance of 1 metre from each other, produces between those two conductors a force of 2×10^{-7} newton per metre length.

EXAMPLE 13.8

Two long parallel wires carry currents of 3A and 4 A respectively in opposite directions. If the separation between them is 10 cm, find the force exerted by one over the other.

Solution

It can be readily seen that the force is that of mutual repulsion.

The magnitude will be given by

$$F = \mu_o I_1 I_2 / 2\pi r$$

Here, $I_1 = 3A$, $I_2 = 4A$, $r = 0.1m$, $\mu_o = 4\pi \times 10^{-7}$

Hence, $F = \frac{4 \times 10^{-7} \times 3 \times 4}{2 \times 0.1} = 2.4 \times 10^{-5}$ newton m^{-1}

13.11 Torque on a Coil

Consider a rectangular loop of wire, ABCD of length l and breadth b suspended freely in a uniform horizontal magnetic field (Fig. 13.21a). Let the plane of the loop be parallel to the field. If there flows, now, a current I in the loop as indicated, each of the vertical sides, AD and BC will experience a force $F = IBl$. By left hand rule, the force on AD will point into the plane of the paper and on BC it will point out of the plane of the paper. The top view of the transverse section of the loop is shown in (Fig. 13.21 b), where the sign ⊙ indicates the cross section of wire AD with current coming out of the paper and the sign ⊗ the cross section of wire BC with current going into the paper. The small arrows show the forces acting on these wires.

These forces constitute a couple and exert a torque on the loop given by

$$\tau = IBl \times b = IBA \qquad \ldots (13.30)$$

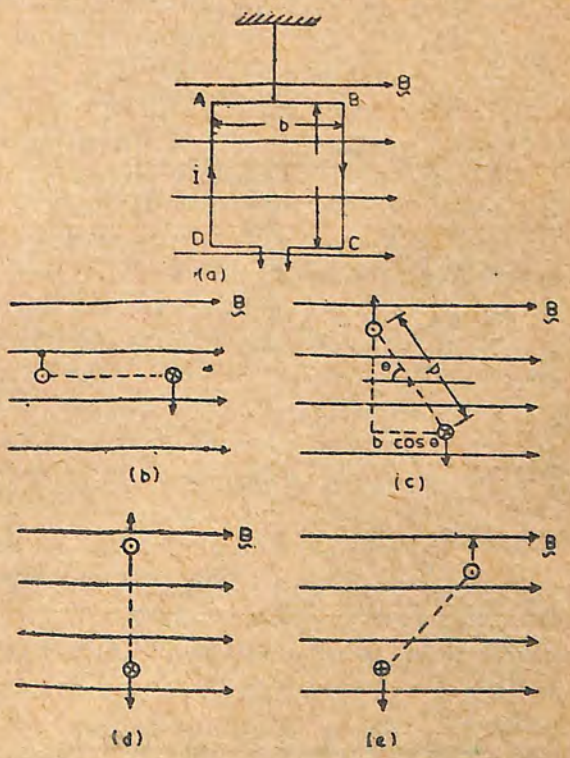


Fig. 13.21 Torque on a freely suspended coil in a uniform magnetic field

where A is the area of the loop. The loop would rotate. As it rotates, the value of torque

changes although the forces on its sides remain the same. When the plane of the loop makes an angle θ with the field (Fig. 13.21 c) the torque is given by

$$\tau = IB\ell \times b \cos\theta = IBA \cos\theta$$

If instead of a single loop of wire, there is a coil of n turns, the torque on the coil is given by

$$\tau = IBAn \cos\theta \quad \ldots (13.31)$$

It can be easily seen that the torque reduces to zero value when the plane of the coil is perpendicular to the magnetic field (Fig. 13.21 d). Regarding the forces on the sides AB and CD, no forces act on them when the coil lies parallel to the field as the direction of current through them is parallel to the field.

It can be shown that the expression (13.31) is valid for any geometrical shape of the coil-circular, rectangular etc., for a given area A of the coil.

Equilibrium position of the coil

If there is no restraint on the coil, it would rotate until it is at right angles to the magnetic field. The torque in this position would be zero but because of its inertia of motion the coil will move past this position. As soon as it crosses $\theta = 90°$ position, a torque in the opposite direction develops (Fig. 13.21 e) which tends to bring it back. The coil would thus oscillate, back and forth, a few times, till its energy is dissipated away in overcoming frictional forces and it would ultimately come to rest at the position $\theta = 90°$.

13.12 The Moving Coil Galvanometer

Consider the arrangement shown in Fig. 13.22. A coil made of many turns of fine copper wire wound over a metal frame is suspended by a torsion fibre in the magnetic field of a permanent magnet. A soft iron cylinder is mounted in the space between the cylinderically curved pole pieces. This makes the field lines radial (Fig. 13.22(b). The coil moves in the clearing between the poles and the cylinder. It can be easily seen that whatever

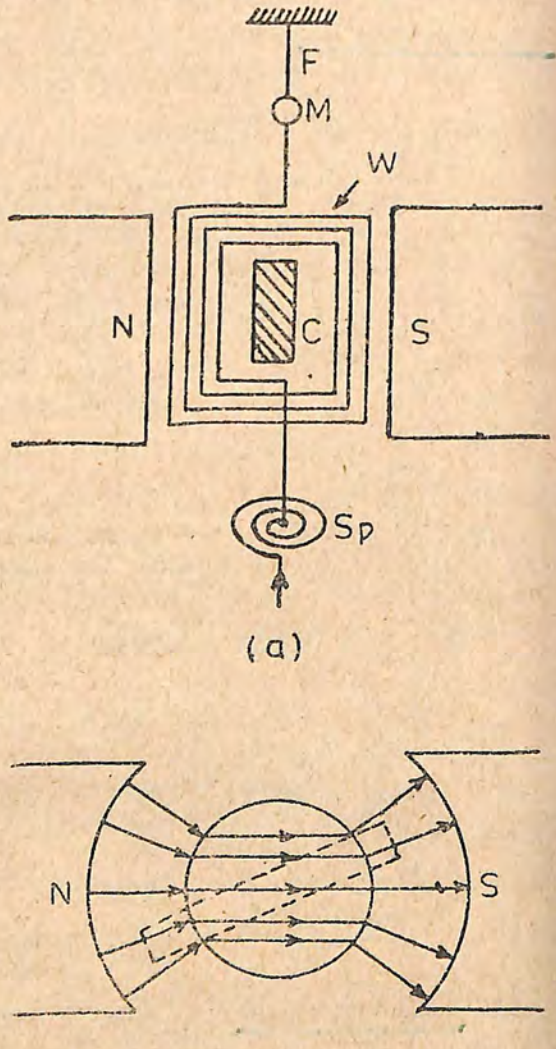


Fig. 13.22 (a) Moving coil galvanometer. F-suspension fibre, M-mirror, W-coil, C-soft iron core, Sp-spring, NS-magnet.
(b) Radial field between pole pieces

be the position of the coil the plane of the coil is always parallel to the field lines. Therefore, the deflecting torque when a current I passes through the coil, using equation (13.31) is given by ($\cos\theta = \cos 0 = 1$)

$$\tau = IBAn$$

where, n is the number of turns on the coil, A is the area of the coil and B is the magnetic field.

This torque will lead to a twist θ in the torsion fibre which sets up a restoring torque, $C\theta$, where C is the torsion constant. C is defined as the torque required to produce a unit twist in the fibre. In the equilibrium position the restoring torque equals the deflecting torque, hence

$$C\theta = IBAn$$

$$\text{or } \theta = \frac{BAn}{C} I \qquad \ldots (13.32)$$

Thus the deflection of the coil is directly proportional to the current in the coil.

The device described above is the basic current measuring device, called the galvanometer. The rotation of the coil is noted by reflection of a narrow beam of light from the small mirror attached to the coil. The suspended coil type instrument is delicate and requires careful handling. Very often a pivoted-coil type (Fig. 13.23) is used which is rugged and portable though less sensitive. The basic principle remains the same.

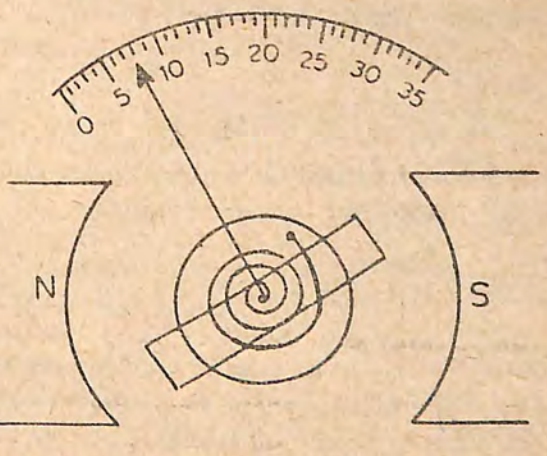


Fig. 13.23 Pivoted-coil type galvanometer

Ammeter

Any instrument used to measure currents is known as an ammeter. The galvanometer described above is a very sensitive ammeter. The pivoted-coil type can measure currents of the order of a microampere. However, the range of measurement of current of a galvanometer is quite small. For example, pivoted coil type, on full scale deflection can measure currents upto a few milliamperes only. It can, however be adopted for measuring large currents in the following manner.

Suppose it is required to measure a current I with the help of a galvanometer which produces a full scale deflection with a current i_g. Let the resistance of galvanometer coil be i_g. A low resistance R_s whose value is small compared with R_g is placed in parallel with the galvanometer as shown in Fig 13.24. This ensures that most of the current passes through

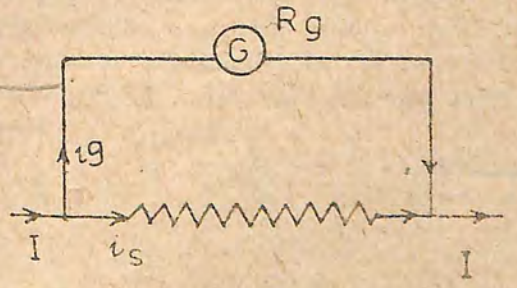


Fig. 13.24 Conversion of a galvanometer into an ammeter

the shunt while the current through the galvanometer does not exceed the safe limit i_g. The current I in the main circuit divides into

two currents, i_g through the galvanometer and i_s through the shunt, then

$$i_g R_g = i_s R_s$$

$$\text{or} \quad R_s = \frac{i_g}{i_s} R_g = \frac{i_g}{I - i_g} R_g \qquad \ldots (13.33)$$

The following example may be illustrative.

EXAMPLE 13.9

A galvanometer whose resistance is 120 ohm has full scale deflection with a current of 0.0005A. Find the value of shunt that must be connected in parallel so that it can read a maximum current of 5A. What is the resistance of the ammeter so constructed ?

Solution

$I = 5A$, $R_g = 120$ ohm

$i_g = 0.0005A$

$$R_s = \frac{i_g}{I - i_g} R_g = \frac{0.0005}{4.9995} 120 = 0.012 \text{ ohm}$$

The resistance of the ammeter is that of the combination R_s and R_g in parallel. It is given by

$$R = \frac{R_s R_g}{R_s + R_g} = \frac{0.012 \times 120}{120 + 0.012} = 0\ 0120 \text{ ohm}$$

A galvanometer, may, therefore be converted into an ammeter to read any current. It is essential that the resistance of the ammeter be very small so that when it is inserted in a circuit it may not alter the value of current it is supposed to measure.

Voltmeter

An instrument used to measure potential difference is known as a voltmeter. A galvanometer can function as a voltmeter. However, the range of measurement of a basic instrument is very small. The maximum potential difference that it can measure is that which will produce the full scale deflection in it and will be given by $V_g = R_g i_g$. Thus, in the above example, the galvanometer can measure a maximum potential difference of $V_g = 120 \times 0\ 0005 = 0.06$ volt only. It can, however, be adapted to measure higher potential differences. For this, it is necessary to connect a high resistance r in series with the galvanometer so that when a potential difference V ($> V_g$) is applied across the combination (Fig. 13.25)

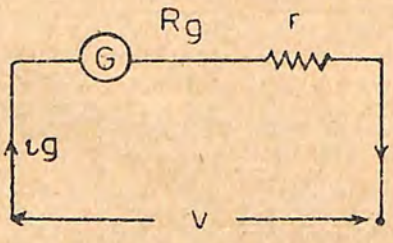


Fig. 13.25 Conversion of a galvanometer into a voltmeter

the current through the galvanometer does not exceed the safe limit i_g. Thus for full scale deflection.

$$V = (R_g + r)\, i_g \qquad \ldots (13.34)$$

The following example may be illustrative.

EXAMPLE 13.10

Find the value of resistance that must be connected in series with the galvanometer of example 13.9 so that it can measure a maximum potential difference of 6 volts.

Solution

The potential difference of 6 volts when applied across the combination of R_g and r, should produce full scale deflection in the galvanometer. Hence

$$6 = (R_g + r)\, i_g = (120 + r)\, 0.0005$$

or, $r = 11880$ ohm

When a voltmeter is used to measure a potential difference V across a resistance R through which a current I is flowing it is desirable that the resistance of the voltmeter should

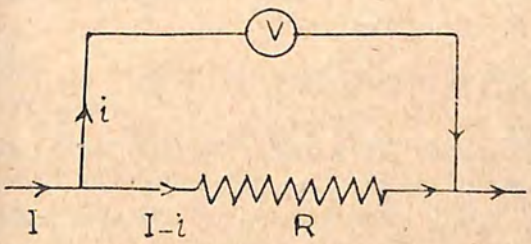


Fig. 13.26 The current through the voltmeter affects the measured value of voltage

far exceed the value of R. The situation is shown in Fig. 13.26. When a voltmeter is connected across R, it draws a current, say, i, leaving a current I-i to flow through R. The potential difference across R now change from $V=R\,i$ to $V'=R(I-i)$. It is V', instead of V, which is measured by the voltmeter. Hence in order that V' may stay close to the desired value V, i should be very small i.e. the voltmeter must have a very high resistance.

Exercises

13.1 Is any work done by a magnetic field on a moving charge ?

13.2 Assuming that the earth's magnetic field is due to a large circular loop of current in the interior of the earth, what is the plane of the loop and what is the direction of current around it ?

13.3 A small coil carrying a current is placed in a uniform magnetic field. How does the coil tend to orient itself relative to the magnetic field ?

13.4 Explain why the two parallel wires carrying current in opposite direction repel each other.

13.5 Two parallel wires carrying current in the same direction attract each other while two beams of electrons travelling in the same direction repel each other Explain why ?
(Hint : A current in the wire produces only a magnetic field while an electron beam is a source of both, an electric and a magnetic field).

13.6 What is the number of free electrons in a piece of silver of area of cross section $1.0\times10^{-4}m^2$ and length 1m. Atomic weight of silver = 108, density of silver $=105\times10^2$ kg/m^3. Assume there is one free electron per atom.
(5.83×10^{24})

13.7 How many electrons pass through a lamp in one minute if the current is 300 mA.
(1.12×10^{20})

13.8 A potential difference V is applied to a conductor of length L, diameter D. How are the electric field E, the drift velocity v_d and the resistance R affected when (i) V is doubled, (ii) L is doubled, (iii) D is doubled.

13.9 A platinum wire has resistance of 10 ohms at 0°C and 20 ohms at 273°C. Find the value of temperature coefficient of resistance. ($\alpha = 1/273$ per K)

13.10 At Rs. 0.40 per kWhr, how much will it cost to have a 60W lamp burning for 5 days ? (Rs. 2.88)

13.11 Near room temperature, the thermo e.m.f. of a copper constantan couple is 40μV per degree. What is the smallest temperature difference that can be detected with a single such couple and a galvanometer of 100 ohm resistance capable of detecting currents as low as 10^{-6} amperes. (2.5°)

13.12 The electron in the hydrogen atom circles around the proton with a speed of 2.18×10^6m/s in an orbit of radius 5.3×10^{-11} m. What magnetic field does it produce at the proton. (12.5 tesla)

13.13 A horizontal wire 0.10m long carries a current of 5A. Find the magnitude and direction of the magnetic field which can support the weight of the wire, assuming its mass is 3.0×10^{-3}kg/m. (5.88×10^{-2} tesla)

13.14 A long straight wire carries a current of 2A. An electron travels with a velocity of 4.0×10^4m/s parallel to the wire 0.1m from it, and in a direction opposite to the current. What force does the magnetic field of current exert on the moving electron ? (2.56×10^{-20} newton)

13.15 The coil of a galvanometer is 0.02×0.08 m. It consists of 200 turns of fine wire and is in a magnetic field of 0.20 tesla. The restoring torque constant of the suspension fibre is 10^{-6} Nm/degree. Assuming the magnetic field to be radial,

(i) What is the maximum current that can be measured by this galvanometer if the scale can accommodate 45° deflection ?

(ii) What is the smallest current that can be detected if the minimum observable deflection is 0.10 degree ? (7×10^{-4} A, 1.6×10^{-6} A)

13.16 A galvanometer has an internal resistance of 1.0 ohm. It gives maximum deflection for a current of 50mA. Show how this instrument can be converted into (i) a voltmeter with a maximum reading of 2.5 volt (ii) an ammeter with a maximum reading of 2.5A.

(series resitance = 49 ohm, shunt resistance = 1/49 ohm)

UNIT 14

Electromagnetic Induction

Earlier we have seen that electric currents generate magnetic fields. In this chapter we shall deal with the converse effect : the production of electric currents with the help of magnetic fields. The phenomenon is known as electromagnetic induction. It was discovered by Faraday in 1831.

Faraday described the phenomena in terms of magnetic flux, which we define first.

14.1 Magnetic Flux

Consider a closed curve, Fig. 14.1, bounding a plane surface of area A. Suppose there

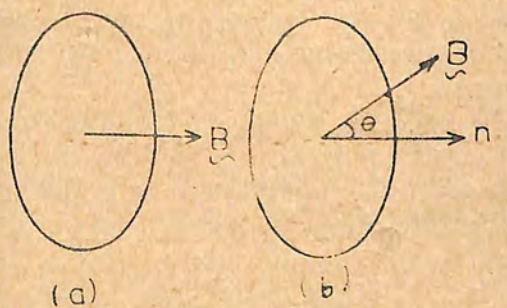


Fig. 14.1 Magnetic flux through an area depends on its orientation w.r.t. the magnetic field

is a uniform magnetic field B perpendicular to the surface. We define the magnetic flux through the area A as

$$\phi = \mathbf{A}.\mathbf{B}. \qquad \ldots 1(4.1)$$

If B is not perpendicular to the surface, the flux through A is given by

$$\phi = \mathbf{B}.\mathbf{A} = BA\cos\theta \qquad \ldots (14.2)$$

where θ is the angle which the vector B makes with the normal to the surface.

In S. I. system, unit of magnetic flux is weber. A uniform magnetic field of 1 tesla, normal to an area of 1 m^2, would give rise to a flux of 1 weber.

Since $B = \phi/A$, field B is quite often expressed as weber/m^2 instead of tesla.

Positive and negative flux

A normal can be drawn to a plane in two ways. The flux is taken as positive when the normal points out in the direction of the field, i.e. $\theta = 0$. When the normal points in the opposite direction, $\theta = 180°$ and flux $= BA \cos 180° = -BA$, that is, flux is negative.

14.2 Laws of Electromagnetic Induction

There are two laws governing the phenomenon of electromagnetic induction. (*i*) Faraday's law, that gives us the magnitude of induced e.m.f. (*ii*) Lenz's law which tells us the direction in which an induced e.m.f. acts

(i) Faraday's Law : Faraday showed that an e.m.f. is induced in a circuit whenever there is a change in the magnetic flux enclosed by a circuit. The magnitude of the induced e.m.f. is equal to the rate of change of flux through the circuit.

Thus

$$|E| = \frac{d\phi}{dt} \qquad \ldots (14.3)$$

(ii) Lenz's Law : The polarity (or sense) of an induced e.m.f. is such that the induced current flows in a direction so as to oppose the change which is causing the induction. Thus, if the magnetic flux is decreasing, the induced current in the circuit must flow in a direction to create flux in the direction of original flux. And, if the flux is increasing the induced current would create a flux that will be opposite in direction to the original magnetic flux.

Remembering Lenz's law, Faraday's law is written as

$$E = -\frac{d\phi}{dt} \qquad \ldots (14.4)$$

When the equation is applied to a coil of N turns, an e.m.f. develops in every turn and the total e.m.f. that appears at the terminals of the coil is the sum of all these e.m.f.s. In such a case, induced e.m.f. is given by

$$E = -N\frac{d\phi}{dt} \qquad \ldots (14.5)$$

If flux changes from ϕ_1 to ϕ_2 in time t, then the average e.m.f. induced is given by

$$E = -N\frac{(\phi_2 - \phi_1)}{t} \qquad \ldots (14.6)$$

If ϕ is in webers and t in seconds, then E is in volts.

EXAMPLE 14.1

A 100 turn close-packed coil of diameter 0.20 m is placed with its plane perpendicular to a uniform magnetic field. The field value, then varies at a uniform rate from 0.10 weber/m^2 to 0.30 weber/m^2 in 5.0×10^{-2}s. Find the e.m.f. induced in the coil.

Solution

$$E = -N\frac{(\phi_2 - \phi_1)}{t}$$

Here $N = 100$

$\phi_2 - \phi_1 = A(B_2 - B_1) = \pi \times 0.10^2 (0.30 - 0.10)$ webers.

$t = 5.0\times10^{-2}$ s,

$$E = -\frac{100\times0.10^2\times0.20\pi}{5\times10^{-2}}$$

$= -12.6$ volts.

Minus sign indicates that the e.m.f. acts in such a direction as to send a current which would oppose the increase in flux.

EXAMPLE 14.2

A coil lies in a uniform magnetic field B with its plane perpendicular to the field, Fig. 14.2(a). In this position the normal to the coil makes an angle 0° with the field. The coil rotates at a uniform rate to complete one rotation in time T. Find the average induced e.m.f. in the coil during the interval when the coil rotates,

(*i*) from 0° to 90° position, Fig. 14.2(b),
(*ii*) from 90° to 180° position, Fig. 14.2(c),
(*iii*) from 180° to 270° position, Fig. 14.2(d),
(*iv*) from 270° to 360° position, Fig. 14.2(a).

Solution

(*i*) initial flux in position (a), $\phi_1 = B\times A$

final flux in position (b),

$\phi_2 = BA\cos 90° = 0$

time $t = \frac{T}{4}$

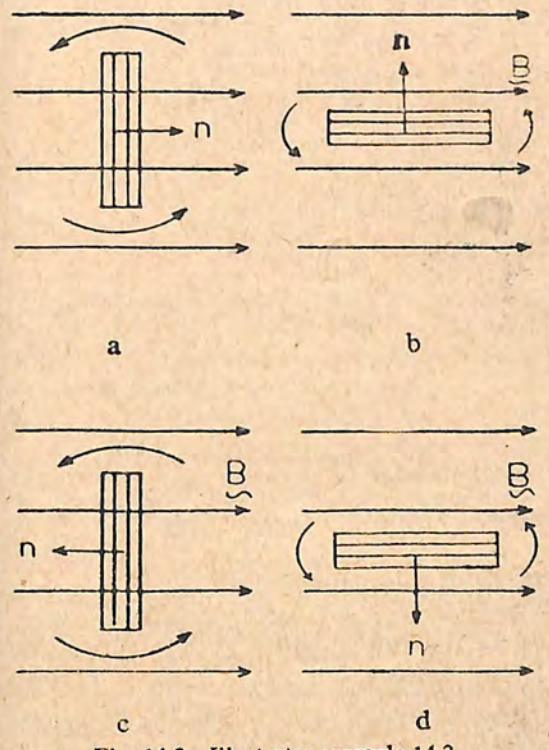


Fig. 14.2 Illustrates example 14.2

Average induced e.m.f.

$$E = -(\phi_2 - \phi_1)/t$$

$$= -\frac{(0 - BA)}{T/4} = \frac{4BA}{T}$$

(*ii*) initial flux, $\phi_1 = 0$

final flux, $\phi_2 = B \times A \cos 180°$

$$= -BA$$

Average induced e.m.f.

$$E = -\frac{(-BA - 0)}{T/4} = \frac{4BA}{T}$$

(*iii*) initial flux, $\phi_1 = -BA$

final flux, $\phi_2 = 0$

Average induced e.m.f.

$$E = -\frac{0-(-BA)}{T/4}$$

$$= \frac{-4BA}{T}$$

(*iv*) initial flux, $\phi_1 = 0$

final flux, $\phi_2 = BA$

Average induced e.m.f.

$$E = -\frac{(BA-O)}{T/4} = \frac{-4BA}{T}$$

Since the sense of the induced e.m.f. in the second half rotation is opposite to that in the first half rotation, the induced current will change its direction after first half rotation.

14.3 Methods of Creating Induced E.M.F.

An induced e.m.f. is produced by the change in magnetic flux passing through a circuit. The magnetic flux may be changed, in view of equation (14.2), by one of the following methods :

(*a*) changing the magnetic field B

(*b*) changing the area A of the loop (circuit)

(*c*) changing the relative orientation of B and A.

Induced e.m.f. by changing B

This can be accomplished in following ways.

(*i*) In Fig. 14.3 an induced e.m.f. will be set up in the wire loop when there is a relative motion between the loop and the magnet. By keeping one of them, the loop or the magnet, fixed and moving the other one changes the magnetic field at the loop. This causes a change in the flux linked with the loop which produces the induced e.m.f. in the circuit. The direction of induced current follows Lenz's law. Thus, in a given situation the induced current flows in the indicated direction so as to oppose the change in magnetic flux through the loop caused by the relative motion. Replacing the magnet with a coil carrying current would give us the same results.

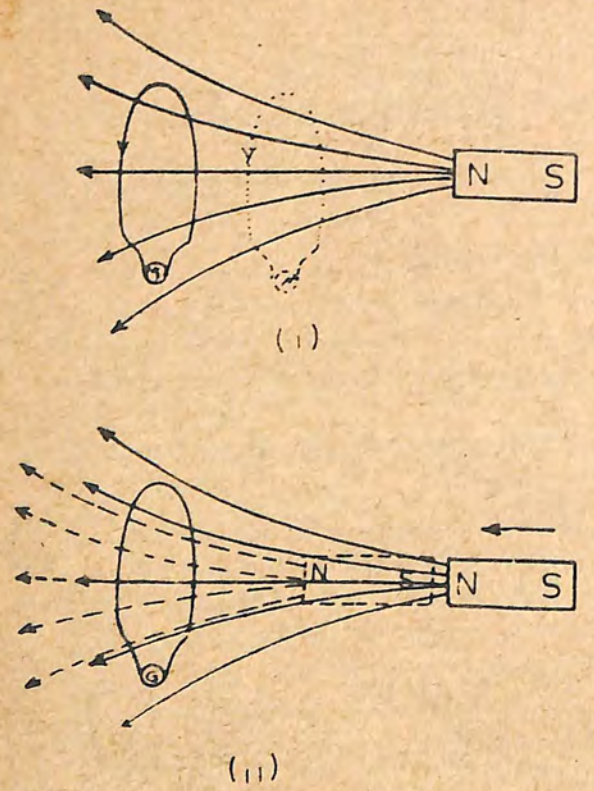


Fig. 14.3 Induced e.m.f. is caused in the loop by change in flux due to change in B at the loop by relative motion of the loop and the source of the magnetic field

(*ii*) Another way of changing B is illustrated in Fig. 14.4. A change in the current in the circuit B produces a corresponding change in the magnetic field at the site of a neighbouring circuit A. The consequent change in the magnetic flux through the loop A causes an induced current to flow through it. The effect is more pronounced when the current in the loop B is switched on or off than when it is varied by sliding the contact in the variable resistor R. The reason is, the magnitude of the induced e.m.f. depends upon the rate of change of flux,

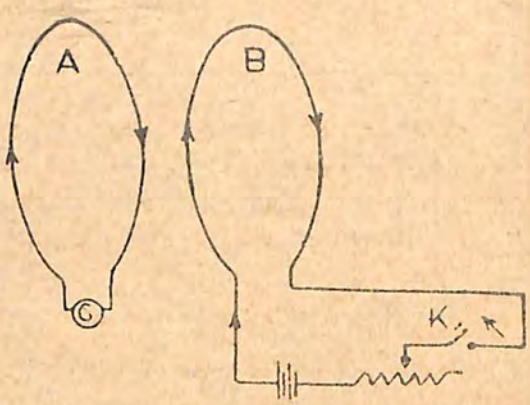


Fig. 14.4 Induced e.m.f. caused by change in the magnetic flux due to changing current in the neighbouring circuit

the faster the rate of change, the greater is the induced e.m.f. and the current caused by it.

Induced e m.f. by changing A

Consider a conductor ab moving with a velocity v towards right on U shaped conducting rails situated in a magnetic field, as shown in Fig. 14.5. The field is uniform and points

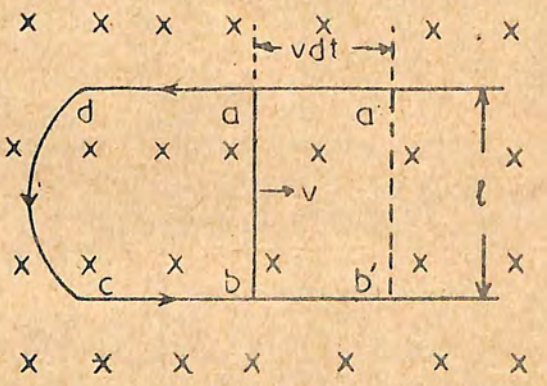


Fig. 14.5 Induced e.m.f. caused by change in flux due to change in area of the circuit

in the plane of the paper. As the conductor slides, the area of the circuit changes from abcd to a' b' c d in time dt. This causes an

increase in flux given by

$$d\phi = B \text{ (area a' b' c d — area abcd)}$$
$$= B.\ \text{area a' b' b a}$$
$$= B\, l\, v\, dt \qquad \ldots (14.7)$$

This gives rise to an induced e.m.f. in the circuit whose magnitude is given by

$$|E| = \frac{d\phi}{dt} = Blv \qquad \ldots (14.8)$$

The induced current flows in the direction shown according to Lenz's law.

EXAMPLE 14.3

An aeroplane with a wingspan of 30 metres flies at a horizontal speed of 100 metres per second in a region where the vertical component of the magnetic field due to earth is 5.0×10^{-5} weber/m^2. What is the potential difference between the tips of the wings ?

Solution

The metal between the wing tips can be considered to be a single conductor which is moving at right angles to a magnetic field. The induced e.m.f. creates a potential difference between the tips. Its value is given by

$$E = Blv$$

Here

$$B = 5.0\times10^{-5} \text{ weber/m}^2,\ l = 30 \text{ metre},\ v = 100 \text{m/s}.$$

Hence,

$$E = 5.0\times10^{-5}\times30\times100$$
$$= 0.15 \text{ volt}$$

Induced e.m.f. by changing relative orientation of the coil and the field

Consider a coil free to rotate about an axis in its own plane, the axis being perpendicular to the magnetic field, as in Fig. 14.6. The flux through the coil when its normal makes an angle θ with the field is given by

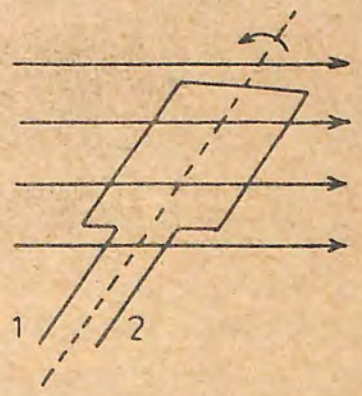


Fig. 14.6 Rotating coil in a magnetic field

$$\phi = BA\cos\theta \qquad \ldots (14.9)$$

As the coil rotates, the magnetic flux through it changes continually. Since the flux is changing there will be an e.m.f. induced in the coil. The direction of the e.m.f. will, however, not remain constant but go on reversing after every half rotation (Example 14.2). The outer terminals 1 and 2 of the coil will alternately become + ve and − ve after intervals of time $\frac{T}{2}$ where T is the period of rotation. If these terminals are connected to an outside circuit, a current will flow in it which will reverse its direction at intervals of time $\frac{T}{2}$.

Let the angular velocity of the coil be ω. If we measure time t from the instant when the coil was perpendicular to the field *i.e.* $\theta = 0$ at $t = 0$ then at time t, $\theta = \omega t$

Therefore, from equation (14.9),

$$\phi = BA\cos\omega t$$

and $$\frac{d\phi}{dt} = -BA\omega\sin\omega t \qquad \ldots (14.10)$$

If the coil consists of N turns, the induced e.m.f. is given by

$$E = -N\frac{d\phi}{dt} = BA\,\omega\,N \sin \omega t \quad \ldots (14.11)$$

The induced e.m.f. is an alternating one. The e.m.f. is maximum when $\omega t = 90°$, *i.e.* when the plane of the coil is parallel to the field. Denoting the maximum value by E_o we have,

$$E_o = BA\omega N \quad \ldots (14.12)$$

The value of E_o depends upon the strength of the field, the area of the coil, the speed of rotation and the number of turns in the coil. E_o is called the amplitude or peak value of e m.f. If the coil makes f rotations per second, we have

$$E = E_o \sin \omega t = E_o \sin 2\pi ft. \quad \ldots (14.13)$$

If E is plotted against t, the graph is a sine curve, as shown in Fig. 14.7(a). Such an e.m.f. is said to be sinusoidal. We also refer to it as alternating e.m.f. This e.m.f. will appear at the terminals of the rotating coil and may be transferred to an external circuit by suitable means.

It may be noted that had we measured the time since the instant when the coil was parallel to the field, *i.e.* at $t = 0$, $\theta = \pi/2$, then at time t we would have, $\phi = BA \cos(\omega t + \pi/2)$. It would have given us, $E = E_o \cos \omega t$. The curve of e.m.f. would have been as shown in Fig. 14.7(b). Except for the matter of choosing the zero of time, there is no essential difference between the sine and cosine forms. Both are known as sinusoidal functions.

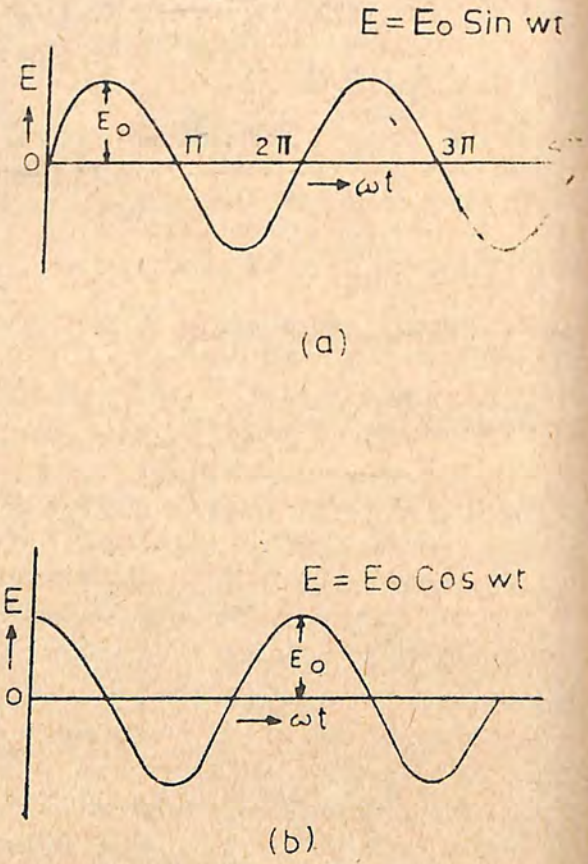


Fig. 14.7 Wave forms of alternating e.m.f.

EXAMPLE 14.4

A rectangular coil of dimensions $0.10\text{m} \times 0.05\text{m}$, consisting of 1000 turns rotates about an axis parallel to its long side, making 3000 revolutions per minute in a field of 100 gauss. What is the maximum e.m.f. induced in the coil, and the instantaneous value of e.m.f. when the coil is at 45° to the field ?

Solution

3000 rev. per minute = 50 rev. per second

$\omega = 2\pi \times 50 = 100\pi$

$E = \omega N AB \sin \omega t$

Here,

$\omega = 100\pi$; $N = 1000$; $A = 0.10 \times 0.05\text{m}^2$;

$B = 100 \times 10^{-4}$ tesla.

$$E = 100\pi \times 1000 \times 0.10 \times 0.05 \times 100 \times 10^{-4} \sin 100\ \pi t$$

$$= 5\pi \sin 100\ \pi t$$

Maximum value of induced e.m.f.,

$$E_0 = 5\pi = 15.7 \text{ volts}$$

When $\omega t = 45°$, $E = E_0 \sin 45°$

$= 15.7 \times \sqrt{\frac{1}{2}} = 11.2$ volts.

14.4 The Generator or Dynamo

What we have described above is the principle of a generator, a means of producing electric current continually by mechanical means. It is the most important method of converting mechanical energy into electrical energy.

Fig. 14.8 shows schematically the essential parts of a generator. A coil abcd capable of

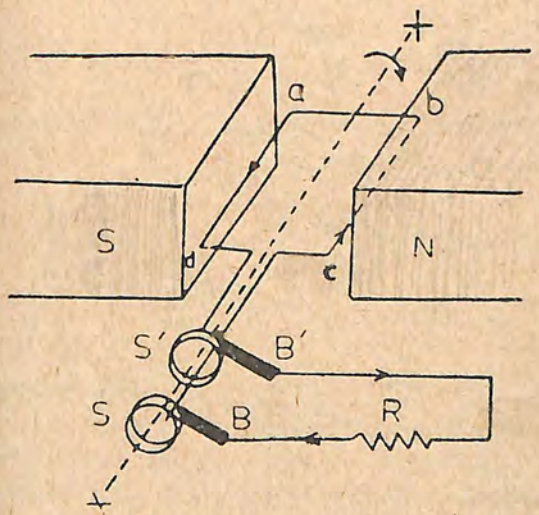


Fig. 14.8 A. C. Generator

rotation about the axis XX is situated in the field of a magnet NS. As the coil rotates an alternating e.m.f. develops in the coil which is fed to the external circuit by means of a pair of metal slip rings, S and S′ which are fixed rigidly to the same shaft which is used to rotate the coil. Each ring is connected to one terminal of the coil and rotates with the coil. The rings maintain sliding contact with the brushes, B and B′. The output e.m.f. of the generator is alternating as shown in Fig. 14.7. As the coil rotates, the current flows out through the brush B for one half of a revolution and through the brush B′ for the next half. The current through the external resistance R is also alternating.

If one directional, *i.e.* direct current is desired, the slip rings are replaced by what is known as a split ring commutator as shown in Fig. 14.9. The commutator rotates with the coil. As before brushes make sliding contact

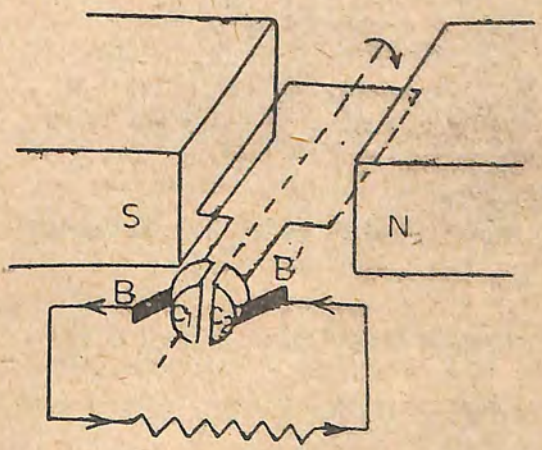


Fig. 14.9 D. C. Generator

with the commutator for half the rotation when C_1 is positive, it is in contact with the brush B, for the other half of rotation it is C_2 which is positive and is in contact with B. Thus the current always leaves the generator through the brush B and we get a direct current in the external circuit. The output, however, is pulsating as shown in Fig. 14.10. If we have several coils, uniformally spaced, as in Fig. 14.11, all

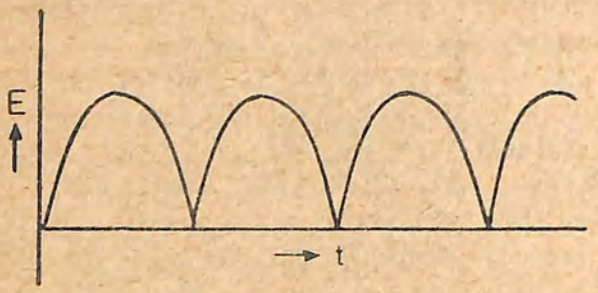


Fig. 14.10 Output e.m.f. from a single coil D. C. Generator

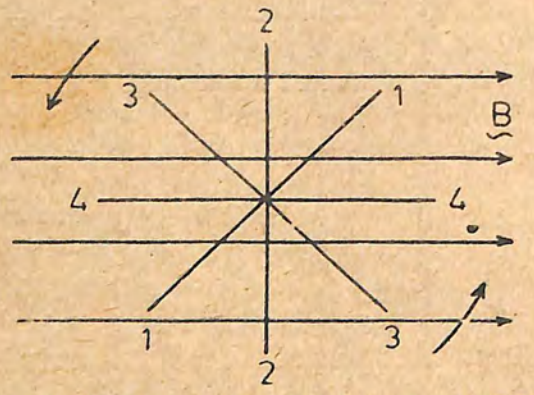


Fig. 14.11 Four coils rotating in a magnetic field

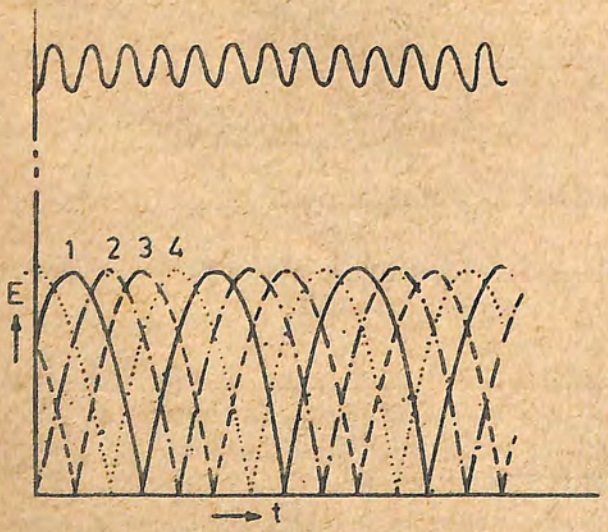


Fig. 14.12 Output e.m.f. from four coils rotating in a magnetic field

connected in series, the maximum value of e.m.f. occurs in each coil at different instant and the net effect is an almost constant unit directional e.m.f. with a small ripple. Fig. 14.12 shows results achieved with four coils. More coils would give an even more smooth output.

14.5 Mutual Inductance

Consider two coils P and S placed near each other, Fig. 14.13(a). If coil P carries a current, it produces a magnetic field which produces a magnetic flux through S. If the current in the coil P, called the primary, is changed, it would cause an induced e.m.f. in the coil S, called the secondary.

Let the current through primary coil at any instant be I_1. Then the magnetic flux at any part of the secondary coil will be proportional to I_1 that is

$$\phi_2 \propto I_1$$

Therefore, the induced e.m.f. in the secondary, when I_1 changes, is given by

$$E = -\frac{d\phi_2}{dt}$$

$$i.e. \quad E \propto -\frac{dI_1}{dt} = -\frac{MdI_1}{dt} \qquad \ldots (14.14)$$

where M is the constant of proportionality, called the mutual inductance of the two coils. It is defined as the e.m.f. induced in the secondary by a unit rate of change of current in the primary. When E is expressed in volts, and $\frac{dI_1}{dt}$ in amperes/second, M is in volt/second/ampere, called henry (H). The mutual inductance between the circuits is 1H, if a rate of change of current of 1A/s in the primary produced an induced e.m.f. of 1 volt in the secondary.

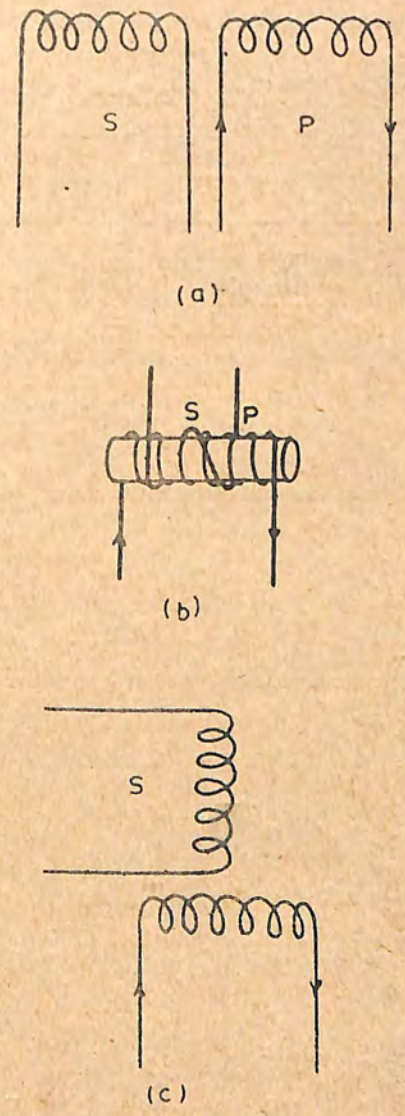


Fig. 14.13 (a) Mutual induction. Variation of current in primary coil P induces an e.m.f. in the secondary coil S.
(b) case of largest value of M
(c) case of least value of M.

M depends upon the number of turns in the coils, their geometrical shape and their separation It is maximum when the entire flux of the primary links with the secondary. Thus the value of M will be minimum when one coil is perpendicular to another as in Fig. 14.13(c). The mutual inductance is further increased, if the coils in Fig. 14.13(b) are wound over an iron core, by a factor μ, where μ is the permeability of iron.

14.6 Self-inductance

When a current flows in a coil, it gives rise to a magnetic flux through the coil itself. If the current strength changes, the flux changes and an e.m.f. is induced in the coil. This e.m.f. is called self-induced e.m.f. and the phenomenon is known as self-induction.

It is easy to see that the flux through the coil is proportional to the current through it, *i. e.*

$$\phi \propto I$$

and, therefore, the induced e.m.f. E is given by

$$E = \frac{-d\phi}{dt}$$

$$i.e.\ E \propto -\frac{dI}{dt} = -L\frac{dI}{dt} \qquad \ldots (14.15)$$

where L is called the self-inductance of the coil. Like M, L is also measured in henry. A coil has self-inductance of 1 henry if an e.m.f. of 1 volt is produced in it when the current passing through it changes at the rate of one ampere per second.

Self-inductance, often called inductance, is a constant of the coil. It depends upon the number of turns, area of cross section, and the permeability of the core material. The larger the number of turns and area of cross section of a coil, the larger is its inductance. If coil is wound over an iron core, the self-inductance

increases by a factor μ where μ is the permeability of iron. A coil possessing an appreciable inductance is known as an inductor.

EXAMPLE 14.5

What e.m.f. will be induced in a 10H inductor in which the current changes from 10 amperes to 7 amperes in 9.0×10^{-2} seconds ?

Solution

$$E = -L\frac{dI}{dt}$$

$$= -L\frac{(I_2 - I_1)}{t}$$

$$= -10\frac{(7-10)}{9.0\times 10^{-2}}$$

$$= 333 \text{ volts}$$

Inductance in a d.c. circuit

Figure 14.14 shows an inductance L and a resistance R (including the resistance of the

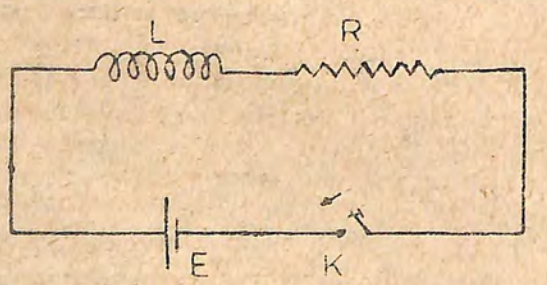


Fig. 14.14 Inductance in a d.c. circuit

coil L) connected to a battery of e.m.f. E. When the key K is closed, the current begins to grow. As the current increases, the magnetic field associated with it also increases and so does the magnetic flux through the coil. This increasing flux induces an e.m.f which produces a current opposed to the one which is growing, (Lenz's law). This limits the rate of rise of the current and it takes some time before the current reaches a steady value given by Ohm's law, $I_0 = E/R$. The effect of inductance, therefore, is to increase the time taken by the current to reach its limiting value I_0. The length of this time depends on the value of L and may vary from a few milliseconds to several seconds. The manner in which I attains the value I_0 is shown in Fig. 14.15 for two values of L.

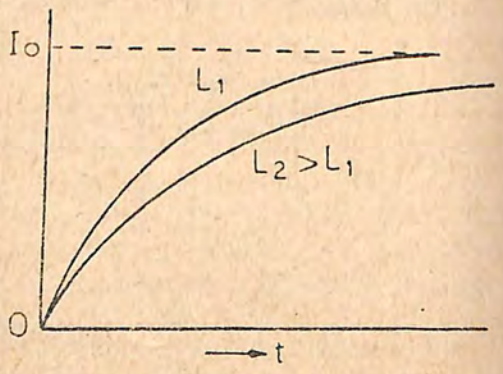


Fig. 14.15 Growth of current in an inductive circuit

Similarly, when a current flowing in a circuit containing inductance is interrupted, an induced e.m.f. is set up in the circuit which tends to maintain the current. Thus in Fig. 14.16 if the switch K is opened, the current decreases rapidly, if this change takes place very rapidly, the induced e.m.f. may be quite

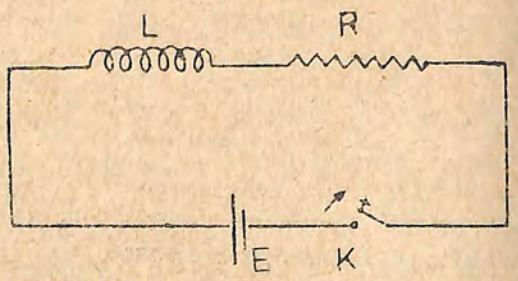


Fig. 14.16 Opening the switch in an inductive circuit leads to sparking

large and may cause a spark to jump across the partially open switch. When the initial current is large and the inductance high, the spark may be sufficiently strong to cause damage to switch contacts and insulation. Switches for d.c. circuits, in such cases, are so designed as to allow the current to decay gradually and thus prevent sparking.

14.7 Some Phenomena Connected with Inductance

(i) Eddy currents

When a sheet of metal is placed in a changing magnetic field, induced currents are set up in the sheet which oppose these changes. These are known as eddy currents. The currents are circular, their sense of flow can be ascertained by applying Lenz's law. Fig. 14.17 shows some eddy currents in a metal sheet placed in an increasing magnetic field pointing in to the plane of paper.

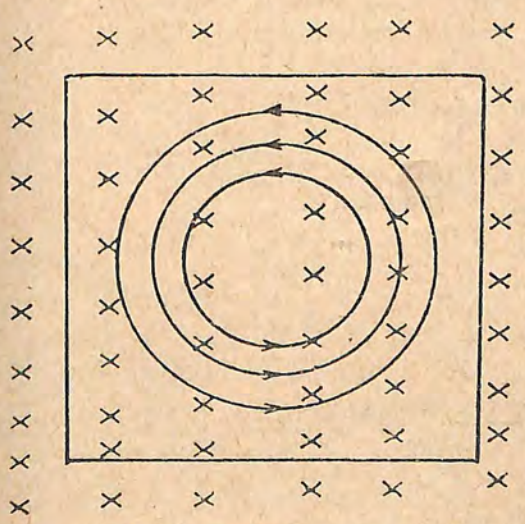


Fig. 14.17 Eddy currents

Since the resistance of metals is quite low, these currents may be quite large and may produce considerable heating effects. Eddy currents are used in heating small metal specimens by putting them in rapidly changing magnetic fields produced by high frequency alternating currents. It is called inducation heating.

Eddy currents are considered undesirable in electrical appliances and machinery, where iron is widely used, as they cause unnecessary heating and wastage of power. This is avoided by building up the desired iron part with thin sheets of the metal which are insulated from each other. The planes of sheets are placed perpendicular to the direction of currents that would be set up by the e.m.f. induced in the material. The insulation between the sheets then offers high resistance to the induced e.m.f. and the eddy currents are substantially reduced.

(ii) Electromagnetic damping

When a current is passed through a galvanometer the galvanometer coil usually suffers a few to and fro oscillations before settling down to its proper deflected position. The motion of the coil is damped largely because of electromagnetic damping—as the coil moves in the magnetic field, a counter e.m.f. is induced in the coil which opposes its motion. The electromagnetic damping can be further increased by winding the coil on a metallic frame. As the frame moves, eddy currents are generated in the frame which use up energy and hence damp the motion. In a properly constructed galvanometer oscillations can be prevented completely—the coil deflects and stays at its final position.

14.8 Alternating Current

Let a source of alternating e.m.f. (represented by the symbol (⁓) be connected across

a resistance R, Fig. 14.18. As the e.m.f., $E = E_0 \sin \omega t$, changes with time, the current

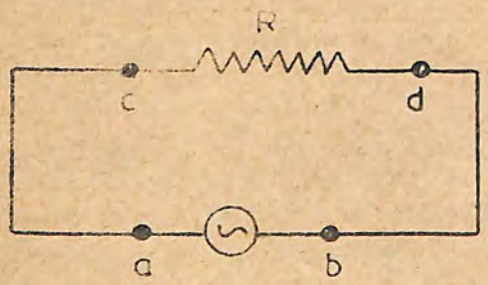


Fig. 14.18 A. C. source with resistance

through the circuit also changes. At any instant, the current would be given by Ohm's law

$$I = \frac{E}{R} = \frac{E_0}{R} \sin \omega t$$

$$= I_0 \sin \omega t = I_0 \sin 2\pi f t. \quad \ldots (14.16)$$

where $I_0 = E_0/R$ is the maximum or peak value of current. f is the frequency and t is the time. The current is sinusoidal and is known as alternating current. Its variation with time

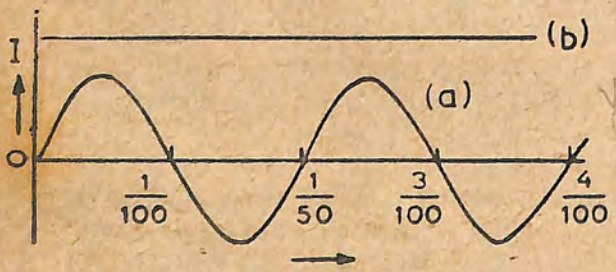


Fig. 14.19 (a) Wave form of alternating current (b) steady current

is shown in Fig. 14.19, the time scale chosen is for the commonly used 50 c/s alternating current. For half the cycle the current flows in one direction and for the other half in the opposite direction. A steady direct current is also shown in the same diagram for comparison.

Since an a.c. current* varies in magnitude continuously and changes direction periodically, the effects produced by it (such as the magnetic effect) will also vary with time. The question immediately arises as to how can we measure such a current ? How do we define an ampere of an alternating current ? There is only one effect which is independent of the direction of the electric current—the heating effect, which we make use of for this purpose. We compare the heating effect of an alternating current with that of a direct current and define the former in terms of the latter.

The effective value of an alternating current, I_{eff} is, therefore, defined as that magnitude of direct current which produces the same heating effect in a given resistance as the given alternating current. Thus an a.c. ampere would be an a.c. current that produces a heating effect equal to that of one d.c. ampere.

It is found that an a.c. current $I = I_0 \sin \omega t$, produces the same heating effect as d.c. current of $I_0/\sqrt{2}$. The value

$$I_{eff} = \frac{I_0}{\sqrt{2}} = 0.707\, I_0 \quad \ldots (14.17)$$

is known as the effective value of the a.c. current. This is also known, sometimes, as root-mean-square or r.m.s. current.†

The effective or r.m.s. value of an a.c. voltage is defined in an exactly similar way and we have

$$E_{eff} = \frac{E_0}{\sqrt{2}} = 0.707\, E_0 \quad \ldots (14.18)$$

* Since a. c. means alternating current, it may seem wrong to use the terms, an a.c. current or an a.c. voltage, from language point of view, but it has now become a common practice to use these terms.

† The reason is, the square root of the mean of the square of the current *i.e.* $\sqrt{(I^2)_{av.}} = I_0/\sqrt{2}$.

Hereafter, whenever numerical values of a.c. voltage and current are given, it is effective values which are meant unless specifically stated otherwise. Thus an a.c. current of 10A means that the current varies in a sinusoidal manner with a peak value of $10\sqrt{2}$ A. An electric supply at 220v means the peak value of voltage is $220\sqrt{2} = 310$v.

14.9 A. C. Circuit Containing Resistance Only

Figure 14.18 shows a resistance R connected in series with an a.c. source of e.m.f.

$$E = E_0 \sin \omega t \qquad \ldots (14.19)$$

At any instant, the potential difference across the resistance R, between the points c and d, is given by

$$V = IR \qquad \ldots (14.20)$$

where I is the instantaneous value of current. At every instant, the p.d. between points c and d must be equal to the p.d. across the source terminals a and b. Thus

$$IR = E_0 \sin \omega t$$

$$\text{or } I = \frac{E_0}{R} \sin \omega t = I_0 \sin \omega t \qquad \ldots (14.21)$$

where $I_0 = E_0/R$ is the amplitude or peak value of the current. Graphs of functions for the instantaneous values, E and I, (Fig. 14.20), show that both the current and the voltage start at zero at the same time, reach maxima at the same time and have the same sinusoidal shape. The voltage and the current under these conditions are said to be in phase.

A voltmeter connected across the resistor would read the effective voltage whose value is $E_0/\sqrt{2}$ Similarly, an ammeter in the circuit would read $I_0/\sqrt{2}$ for the effective current. We,

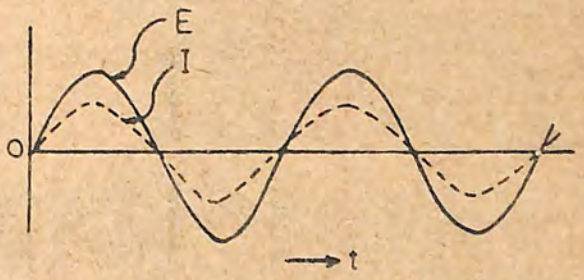


Fig. 14.20 Current is in phase with voltage in a resistive a.c. circuit

then, have

$$I_{eff} = \frac{I_0}{\sqrt{2}} = \frac{E_0}{R\sqrt{2}} = \frac{E_{eff}}{R} \qquad \ldots (14.22)$$

14.10 A.C. Circuit Containing Inductance Only

Although it is difficult in practice to have a pure inductance, it is useful to consider the effects of a pure inductance in an a.c. circuit.

When a steady current flows through a pure inductance, the potential difference between points X and Y (Fig. 14.21) is zero as no resistance is offered to the current. If the current

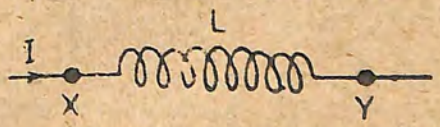


Fig. 14.21 P.D. across an inductor L is given by LdI/dt

is changing then an induced e.m.f. (—LdI/dt) exists in the inductor which opposes the change in the current. This makes the potential of Y different from X. The potential drop, when the current flows from X to Y is given by

$$V_X - V_Y = L\frac{dI}{dt} \qquad \ldots (14.23)$$

Now consider the circuit of Fig. 14.22. Suppose an alternating current given by

$$I = I_0 \sin \omega t \qquad \ldots (14.24)$$

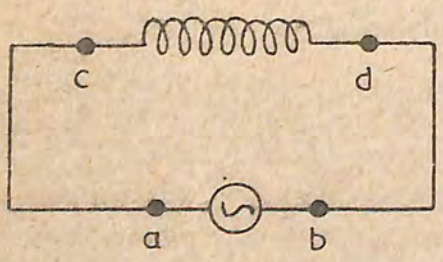


Fig. 14.22 A.C. source with an inductance

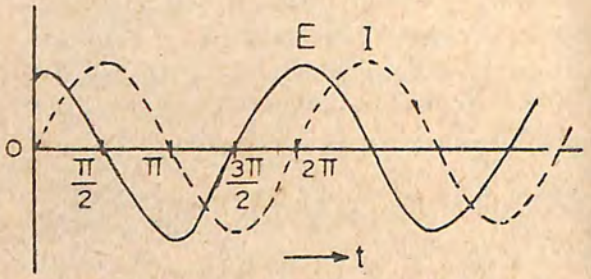


Fig. 14.23 Current I lags the voltage E by $\pi/2$ in an inductive circuit

flows in the inductor. What must be the applied e.m.f. in the circuit in order to produce such a current ?

The instantaneous value of p.d. across L, between points c and d, is given by, equation (14.23),

$$V = L\frac{dI}{dt}$$

$$= L\frac{d}{dt}(I_o \sin \omega t)$$

$$= L\omega I_o \cos \omega t \quad \ldots (14.25)$$

Now, the p.d. between points c and d must, at every instant, be equal to the p.d. across the source terminals a and b. Hence the instantaneous value of applied e.m.f. is given by

$$E = L\omega I_o \cos \omega t = E_o \cos \omega t$$

$$= E_o \sin (\omega t + \pi/2) \quad \ldots (14.26)$$

where $E_o = L\omega I_o$ is the amplitude or peak value of the applied e.m.f. Comparing with the resistive case, $E_o = RI_o$, we find ωL plays the same role here as the resistance R in the earlier case, that is, it impedes the flow of current. The quantity

$$\omega L = X_L = 2\pi fL \quad \ldots (14.27)$$

is called the inductive reactance of the circuit and has the unit of ohm.

From equations (14.24) and (14.26), it is seen that the current and the voltage are 90° out of phase. Fig. 14.23 shows the way in which the instantaneous values, I and E, vary, I reaches maximum value a quarter of cycle later than E. One says that the pure inductance causes the current to lag behind the e.m.f. in phase by $\pi/2$ (radians).

A voltmeter connected across the inductor would read $E_0/\sqrt{2}$ for the effective voltage and an ammeter in the circuit would read $I_o/\sqrt{2}$ for the effective current.

We then have,

$$I_{eff} = \frac{I_o}{\sqrt{2}} = \frac{E_o}{L\omega\sqrt{2}} = \frac{E_{eff}}{L\omega} = \frac{E_{eff}}{X} \quad \ldots (14.28)$$

Inductive reactance X_L plays the same role here as the resistance R in equation (14.22).

EXAMPLE 14.6

The current through a 1.0 henry inductor varies sinusoidally with an amplitude of 0.5 amperes and a frequency of 50 cycles per second. Calculate potential difference across the terminals of the inductor.

Solution

$I = I_o \sin 2\pi ft$

p. d. across the inductor is given by,

$$V = L\frac{dI}{dt} = L\frac{d}{dt}(I_o \sin 2\pi ft)$$

$$= 2\pi fL.I_o \cos 2\pi ft$$

Amplitude of voltage is,

$$2\pi f L I_o = 2 \times 3.14 \times 50 \times 1.0 \times 0.5 = 157 \text{ volts}$$

$$V = 157 \cos 100\pi t$$

Therefore,
The voltmeter will read the effective value

$$V_{eff} = \frac{V_o}{\sqrt{2}} = \frac{157}{\sqrt{2}} = 112 \text{ volts.}$$

EXAMPLE 14.7

What is the inductive reactance of a coil if the current through it is 80 mA and voltage across it is 40 V ?

Solution

These are the effective values hence,

$$X_L = \frac{V_{eff}}{I_{eff}} = \frac{40}{80 \times 10^{-3}} = 500 \text{ ohms.}$$

EXAMPLE 14.8

At what frequency will a 0.5 henry inductor have a reactance of 2000 ohms ?

Solution

$$X_L = \omega L = 2\pi f L$$
$$2000 = 2\pi f \times 0.5$$
$$f = \frac{2000}{\pi} = 637, H_z$$

14.11. A. C. Circuit Containing Capacitance Only

A capacitor consists of two plates of conducting material separated by an insulator. Its resistance is, therefore, practically, infinite. We might expect that when a capacitor is put into a circuit, d. c. or a. c., no current would flow. This is not so. Let us first consider a capacitor in a d.c. circuit.

Capacitor in a d.c. circuit : Consider the circuit shown in Fig. 14.24. As soon as we

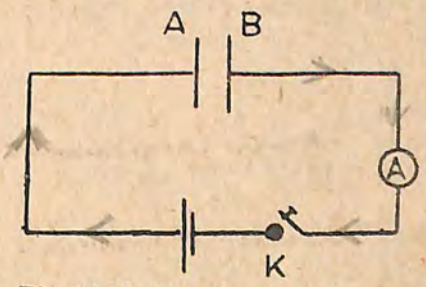


Fig. 14.24 Capacitor in a d.c. circuit

press the key, electrons begin to flow from the negative terminal of the battery to the plate B and from plate A to the positive terminal of the battery. The plate A thus begins to acquire positive charge and the plate B, negative charge. This charging of the capacitor continues till the potential difference between the plates becomes equal to that across the battery terminals. Then, it ceases. This flow of charge is equivalent to a current. Thus a current does flow in the circuit, during the charging process; though not through the capacitor but in the remainder of the circuit. The ammeter would, therefore, show a momentary deflection. The direction of current is from the plate B to the plate A via the battery. Its magnitude at any instant is given by the rate of growth of charge on the capacitor.

$$I = \frac{dQ}{dt} \qquad \ldots (14.29)$$

The charging process can be extended in time if we include a resistance in the circuit, Fig. 14.25. It is found that the final charge on the capacitor, given by

$$Q_o = V_o C \qquad \ldots (14.30)$$

where, V_o = battery voltage, C = capicitance, is reached asymptotically as shown in Fig. 14.26. (Compare with the growth of current in a d.c. inductive circuit).

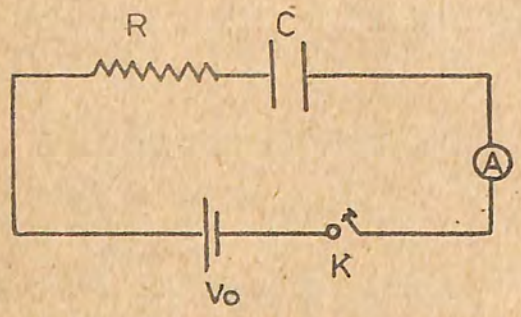


Fig. 14.25 Charging of a capacitor through a resistance

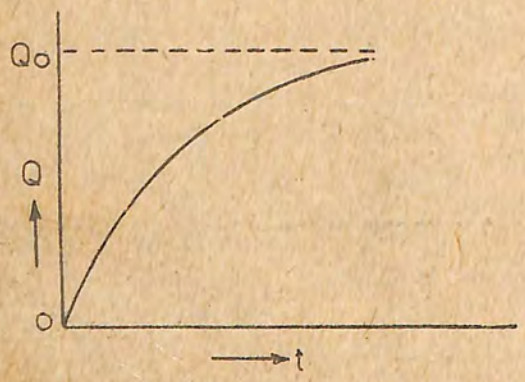


Fig. 14.26 Growth of charge on capacitor with time

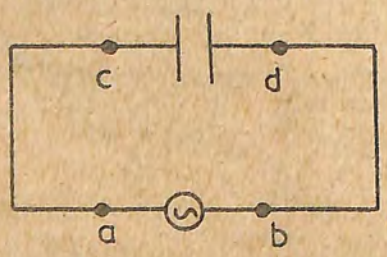


Fig. 14.27 A. C. source with a capacitor

Capacitor in an a.c. circuit: Now consider the capacitor in an a.c. circuit, Fig. 14.27. As the voltage from the source is continually changing, the charge on the capacitor is also continually changing. During a complete cycle, the capacitor is first charged in one direction, then discharged, again charged in the reverse direction and discharged. As this charging and discharging of the capacitor is taking place continuously, a continuous current exists in the circuit. Let us find out the nature of this current.

It is evident that the p.d. across the capacitor between points c and d, at every instant, has to be exactly the same as that across the source terminals a and b. Therefore the capacitor must charge and discharge in such a manner that the p.d. across it, V, is sinusoidal and equal to the applied e.m.f. at every instant. That is,

$$V = E = E_o \sin \omega t$$

The charge on the capacitor, at any instant, is given by

$$Q = CV$$

Therefore, current at any instant is given by

$$I = \frac{dQ}{dt} = C\frac{dV}{dt} = C\frac{d}{dt}(E_o \sin \omega t)$$
$$= E_o C\omega \cos \omega t$$
$$= I_o \cos \omega t$$
$$= I_o \sin (\omega t + \pi/2) \quad \ldots (14.31)$$

where,

$$I_o = \frac{E_o}{1/C\omega} \quad \ldots (14.32)$$

Thus, in this case, the current is sinusoidal but 90° ahead of e.m.f., in phase. The wave form of E and I are shown in Fig. 14.28. The quantity

$$X_o = \frac{1}{C\omega} = \frac{1}{2\pi fC} \quad \ldots (14.33)$$

is known as the capacitive reactance of the circuit. It plays the same role in this case, as the inductive reactance X_L in the inductive case. Its unit is ohm.

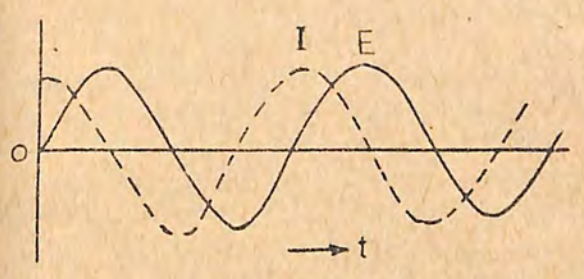


Fig. 14.28 Current I leads the voltage E by $\pi/2$ in a capacitive circuit

A voltmeter connected across the capacitor would read $E_0/\sqrt{2}$ for the effective voltage and an ammeter in the circuit would read $I_0/\sqrt{2}$ for the effective current. We, then, have

$$I_{eff} = \frac{I_0}{\sqrt{2}} = \frac{E_0}{\sqrt{2}(1/C\omega)} = \frac{E_{eff}}{1/C\omega} = \frac{E_{eff}}{X_C} \quad \ldots (14.34)$$

EXAMPLE 14.9

What is the capacitive reactance of a 5-μf capacitor when it is part of a circuit whose frequency is (*i*) 50c/s (*ii*) 10^6c/s ?

Solution

$$(i)\ X_C = \frac{1}{2\pi \times 50 \times 5 \times 10^{-6}} \text{ ohms}$$

$$= 637 \text{ ohms}$$

$$(ii)\ X_C = \frac{1}{2\pi \times 10^6 \times 5 \times 10^{-6}}$$

$$= 3.18 \times 10^{-2} \text{ ohms.}$$

14.12 LCR Circuit

A resistor, an inductor, or a capacitor, each of these circuit elements impedes an alternating current. The impeding effect is measured by the resistance (R), the inductive reactance (X_L) and the capacitive reactance (X_C), respectively, which are defined by

$$\frac{V}{I} = R \text{ (for resistor)}$$

$$\frac{V}{I} = X_L \text{ (for inductor)}$$

$$\frac{V}{I} = X_C \text{ (for capacitor)}$$

where V is the effective voltage across the circuit element and I the effective current through it.

Where a combination of these elements forms part of a circuit the total current impeding effect is defined, in an analogous manner by,

$$Z = \frac{V}{I} \quad \ldots (14.35)$$

where Z is known as the impedance of the combination. Its unit, obviously, is ohm.

In Fig. 14.29 we have a series combination of a resistor of resistance R, an inductor of reactance X_L and a capacitor of reactance X_C. It turns out that the total impedance Z of the combination is given by

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \quad \ldots (14.36)$$

and not by $Z = R + X_L + X_C$. The reason for this peculiar relationship is the fact that the reactances (X_L, X_C) arise because of a somewhat different physical mechanism than ordinary resistance (R). We cannot, therefore, combine them by simple addition. Equation (14.36) is a direct consequence of the vectorial addition of effective voltages in an a.c. circuit as discussed below.

In the circuit of Fig. 14.29, at any instant, the applied voltage is equal to the sum of the voltage drops across individual elements.

$$V = V_R + V_L + V_C \text{ (instantaneous values)}$$

However, if we measure the effective voltages

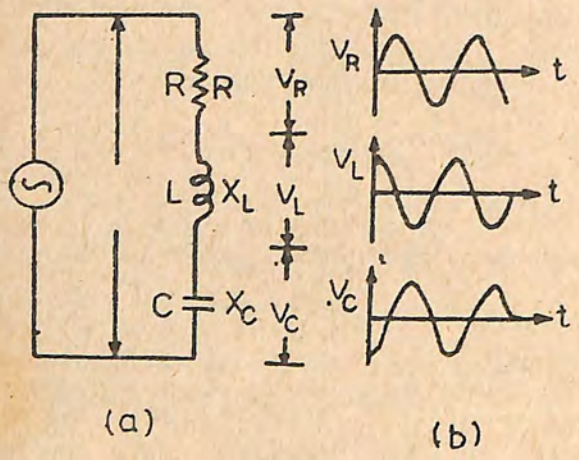


Fig. 14.29 An LCR series circuit

(with a voltmeter) we would find that

$$V \neq V_R + V_L + V_C \text{ (effective values)}$$

The reason is, although same current passes through each component, the voltage across each of them bears a different phase relationship with the current and so the voltages are out of phase with one another. Mathematical analysis shows that effective voltages (and currents) add up vectorially. It is customary to treat them as vectors for the purpose of circuit analysis. Phase differences are represented by angles between the vectors.

Thus, in Fig. 14.30, the current I is represented by the vector OI; V_R, the voltage across R, being in phase with the current, is represented by vector OR; V_L, the voltage across L, being ahead of current by $\pi/2$, by vector OL and V_C, the voltage across C, since it lags behind the current by $\pi/2$, by vector OC. The resultant vector OP represents the effective voltage across the combination. Its magnitude is given by,

$$V = \sqrt{V_R^2 + (V_L - V_C)^2} \quad \ldots (14.37)$$

Since $V_R = RI$, $V_L = X_L I$, $V_C = X_C I$ we have

$$V = \sqrt{R^2 I^2 + (X_L - X_C)^2 . I^2}$$

$$= \sqrt{R^2 + (X_L - X_C)^2} . I$$

or $$\frac{V}{I} = Z = \sqrt{R^2 + (X_L - X_C)^2} \quad \ldots (14.38)$$

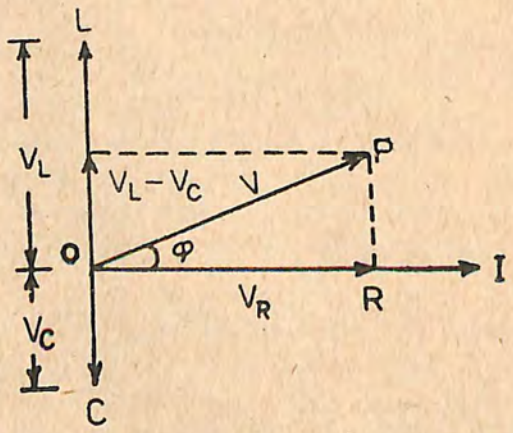


Fig. 14.30 Vectorial representation of voltages in an a.c. circuit

The relationship between impedance (Z), resistance (R) and reactances (X_L, X_C) is illustrated in Fig. 14.31 by constructing a vector diagram similar to Fig. 14.30.

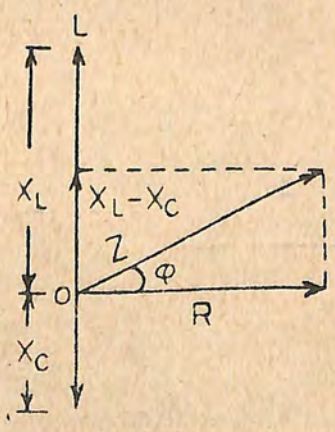


Fig. 14.31 Impedance diagram

It may be seen from Fig. 14.30 that the current I lags behind the voltage V by phase angle ϕ. Thus, the circuit is inductive. This is so because we have assumed $V_L > V_C$ (*i.e.* $X_L > X_C$). If $X_L < X_C$, the combination would be capacitive and the current would lead the voltage. The phase angle, also called the angle of lead or lag, is given by

$$\tan\phi = \frac{X_L - X_C}{R} \quad \text{or} \quad \cos\phi = \frac{R}{Z} \qquad \ldots (14.39)$$

EXAMPLE 14.10

A resistor of 100 ohms, an inductor of 0.5 henry, and a capacitor of 10 microfarads are connected in series. A 220 volt 50 cycle alternating potential is connected across the group. Find (a) the impedance of the circuit, (b) the current, (c) the potential difference across each of the three elements, and (d) the phase angle between the current and the applied voltage. Construct the vector diagram for the voltages.

Solution

(*a*) $X_L = 2\pi fL = 2\pi \times 50 \times 0.5 = 157$ ohm.

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi \times 50 \times 10 \times 10^{-6}}$$

$$= 318.4 \text{ ohm}$$

$$Z = \sqrt{(100)^2 + (157 - 318.4)^2}$$

$$= 189.5 \text{ ohm.}$$

(*b*) $I = \frac{220}{189.5} = 1.16$ ampere

(*c*) Voltage across the resistance $V_R = I \times R = 1.16 \times 100 = 116$ volts

Voltage across the inductance $V_L = I \times X_L = 1.16 \times 157 = 182$ volts

Voltage across the capacitor $V_C = I \times X_C = 1.16 \times 318.4 = 369$ volts

(*d*) $\cos\phi = \frac{R}{Z} = \frac{100}{189.5} = 0.5263$, $\phi = 58.3°$

Vector diagram for voltages is shown in Fig. 14.32

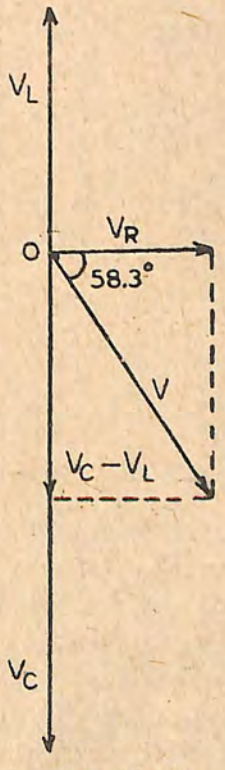


Fig. 14.32

Resonance

It may be seen if $X_L = X_C$, the impedance

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

is at its minimum value and is simply equal to R. The circuit is purely resistive. The current is maximum. The applied voltage and the current are in phase. This is known as the condition of resonance and the frequency f_r at which this occurs is known as resonant frequency. This happens when

$$X_L = X_C$$

or $$2\pi f_r L = \frac{1}{2f_r \pi C}$$

or $$f_r = \frac{1}{2\pi\sqrt{LC}} \quad \text{...(14.40)}$$

14.13 Transformer

One of the most useful applications of electromagnetic induction is the transformer. The construction of a simple transformer is shown in Fig. 14.33. Two coils, each consisting of many turns of wire are wound on a

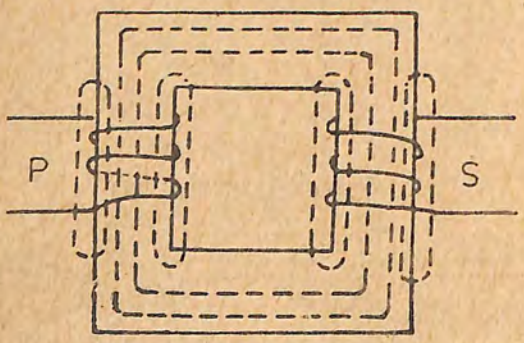


Fig. 14.33 Transformer : P-Primary, S-Secondary

continuous iron core. One of these coils called the 'primary' is connected to an a.c. source. The other coil, called the 'secondary' is connected to the 'load' which may be a resistance or any other electrical device to which electric power is to be supplied.

The alternating current in the primary produces an alternating magnetic flux in the core which passes through the secondary coil also. As there is very little 'leakage' the magnetic flux through the secondary is almost the same as through the primary. This changing magnetic flux produces an induced e.m.f. in the secondary and it also causes a self-induced back e.m.f. in the primary.

Consider the situation when no load is attached to the secondary, that is, its terminals are open. Let N_1 and N_2 be the number of turns in the primary and secondary, respectively. The induced e.m.f. in the primary is given by

$$E_1 = -N_1\frac{d\phi}{dt} \quad \text{... (14.41)}$$

where, I_1 is the current in the primary and ϕ the magnetic flux in the core, at any instant. The induced e.m.f. in the secondary is given by

$$E_2 = -N_2\frac{d\phi}{dt} \quad \text{... (14.42)}$$

Thus,

$$\frac{E_2}{E_1} = \frac{N_2}{N_1} \quad \text{... (14.43)}$$

The value of induced e.m.f. in the primary can be shown to be almost equal to the applied e.m.f.

Let the applied e.m.f. to the primary be $E = E_0 \sin \omega t$. In the primary circuit, the self induced e.m.f. E_1 acts opposite to the applied e.m.f. (Lenz's law). At any instant it is the difference, $E - E_1$, which sends the current I_1 through the resistance R of the primary,

Thus,

$$E - E_1 = RI_1$$

However, R is very small, and so the term RI_1 can be neglected. Hence

$$E = E_1 \quad \text{... (14.44)}$$

Thus, in equation (14.43) E_1 may be described as the input e.m.f. given to the primary and E_2 as the out-put e.m.f. from the secondary. We have

$$\frac{E_2}{E_1} = \frac{\text{output e.m.f.}}{\text{input e.m.f.}} = \frac{N_2}{N_1} \quad \text{... (14.45)}$$

If V_1 and V_2 be the effective values of e.m.f.s, then

$$\frac{V_2}{V_1} = \frac{N_2}{N_1} \quad \ldots (14.46)$$

The quantity N_2/N_1 is called the 'turns ratio' of the transformer. When $N_2 > N_1$, $V_2 > V_1$ the transformer is known as step-up transformer. If $N_2 < N_1$, $V_2 < V_1$ the transformer is known as step-down transformer.

The law of conservation of energy requires that the energy delivered to the secondary circuit must be equal to or less than the energy supplied to the primary. Assuming the transformer to be ideal, with no energy losses, the average power input must be equal to the average power output. Therefore

$$V_1 I_1 = V_2 I_2 \quad \ldots (14.47)$$

where the voltages and currents are the effective values. From eq. (14.47) we have

$$\frac{I_2}{I_1} = \frac{V_1}{V_2} = \frac{N_1}{N_2} \quad \ldots (14.48)$$

The current in the secondary decreases in the same ratio as that by which the voltage increases.

The efficiency of a transformer is defined as

$$\eta = \frac{\text{Power output}}{\text{Power input}}$$

In real transformers, the efficiency is fairly high (90-99 per cent) though not 100 per cent. There are several causes for power loss. The main losses are two : (i) the I^2R loss-due to heating of copper wires used in the windings. This can be minimized by using thick wires. (ii) Core loss-due to work done in carrying the iron core through cycles of magnetization and demagnetization. This is minimized by the choice of iron with special magnetic properties. In addition, there is power loss due to eddy currents which is reduced by using laminated iron core (Fig. 14.34). There is also some loss due to flux leakage but it is quite small.

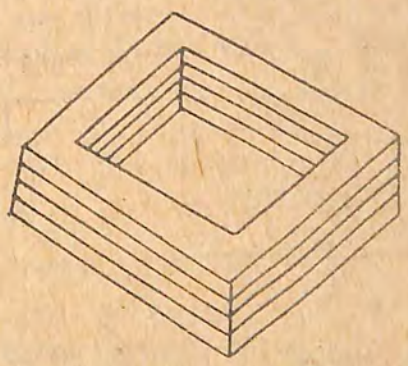


Fig. 14.34 Laminated core

Transmission of electric power

Electric power stations are, generally, situated in remote areas where it is cheaper to produce electric power. This power has to be transmitted to the cities and areas where it is needed. This is done by transmission lines which consist of two parallel wires for carrying current from and to the power station.

To avoid the loss of power due to I^2R losses in the line wire, the output, voltage of the generator is first 'transformed' to a much higher value by a step-up transformer. It converts the electric power at low voltage and high current to the same power at higher voltage and lower current. Due to reduction in the value of current, the I^2R losses in the lines are reduced. To appreciate the economy in transmitting power at high voltage, let us consider the following example.

Suppose a power generator produces 25 KW of power at 100 amperes and 250 volts. It is desired to deliver this power to a consumer 1 Km away on a transmission line whose resistance is, say, 1 ohm. The line loss is given by $I^2R = 100^2 \times 1 = 10{,}000$ W $= 10$ KW. Thus

40 per cent of power would be wasted. If we step-up the voltage by a transformer to 2500 volts, the current in the lines would reduce to 10 amperes (Fig. 14.35). The I^2R loss would come to $10^2 \times 1 = 100W$ which is negligible.

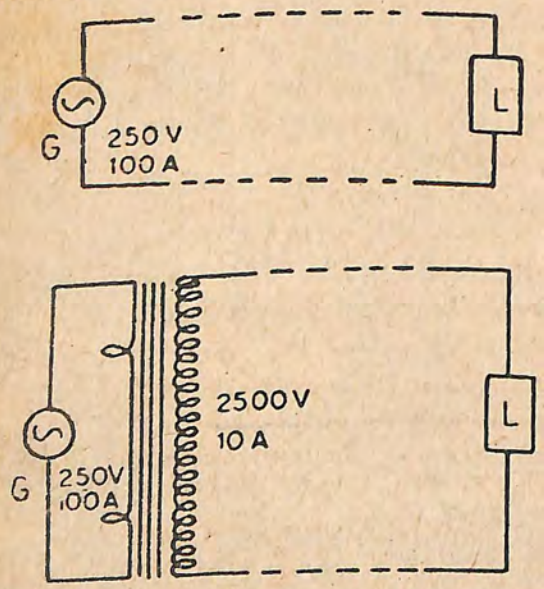


Fig. 14.35 Transmission of power at higher voltage reduces line losses. G - Generator, T - Step-up transformer, L - Load.

A typical power generator gives an output of 1000 KW at 6600 volts. In practice this voltage is stepped up to 132000 volts before transmission. The cables used for transmitting power over long distances are suspended by large porcelain insulators from large steel structures (pylons). The main transmission lines from power stations form part of a common system called the 'grid' which covers a large region of the country. Power from all the power stations in the region is fed into the grid which forms a common pool from which power can be drawn where needed. This allows an efficient power distribution and acts as a safeguard for ensuring a minimum power supply to consumers in the event of failure of power generation at some station. From the grid, the power is fed to the cities at 33000 V, the stepping down is done outside the city. Then again at a sub-station, the supply is stepped down to 6600 volt. Power is supplied to the big consumers like factories at this voltage which they can further step-down according to their needs. For ordinary domestic consumers the voltage is again reduced to 220 V.

Exercises

14.1 What are the dimensions of magnetic flux ?

14.2 Mark the current direction in the secondary windings of Fig. 14.36 as the switch is closed.

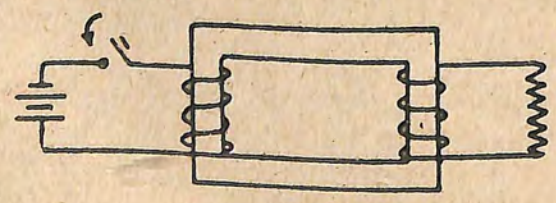


Fig. 14.36

14.3 Show that Lenz's law is a consequence of conservation of energy principle.

14.4 What factors govern the magnitude of the e.m.f. induced in an electric circuit ?
What factors govern the direction of e.m.f. ?

14.5 Find the magnitude of e.m.f. induced in a 200 turn coil with cross-sectional area of 0.16 m^2 if the magnetic field through the coil changes from 0.10 weber/m^2 to 0.50 weber/m^2 at a uniform rate over a period of 0.02 seconds. (640 volts)

14.6 Distinguish between self and mutual induction.

14.7 If the speed of generator is increased, how would it affect (*i*) the maximum e.m.f. produced ? (*ii*) the frequency of the e.m.f. ?

14.8 The electric mains in the house is marked 220 V, 50 c/s. Write down the equation for instantaneous voltage. ($310 \sin 100\pi t$)

14.9 A rectangular coil of 200 turns of wire, 15 cm × 40 cm, makes 50 revolutions/second about an axis perpendicular to the magnetic field of .08 weber/m^2. What is the instantaneous value of induced e.m.f. when the plane of the coil makes an angle with the magnetic field lines of (*i*) 0° (*ii*) 60° (*iii*) 90° ? (301.4V, 150.7V, 0)

14.10 What is the difference between a.c. and d.c. ? What is the advantage of a.c. over d.c. ? Can we use 15c/s a.c. for lighting purposes ?

14.11 What is the significance of a negative value of instantaneous current ?

14.12 Distinguish between resistance, reactance and impedance.

14.13 Sketch graphs to show how the reactance of (*i*) a capacitor, (*ii*) an inductor varies as a function of frequency ?

14.14 What is (i) the reactance of an inductor in a d.c. circuit (ii) the reactance of a capacitor ?

14.15 What is meant by the statement that the current through an inductor lags the e.m.f. across it by $\pi/2$?

14.16 The total impedance of the circuit decreases when a capacitor is added in series with a given impedance. Explain why ?

14.17 If the effective current in a 50-cycle a.c. circuit is 50, what is (*a*) the peak value of current ? (b) the value of current 1/300 second after it was zero ? (7.0A, 6.06A)

14.18 When 100 volts d.c. are applied across a coil, a current at 1 ampere flows through it. When 100 volts a.c. of 50 c/s are applied to the same coil only 0.5 ampere flows. Calculate the resistance, the impedance and the inductance of the coil. (100Ω, 200Ω, 0.55Ω)

14.19 A 25.0μF capacitor, a 0.10-henry inductor and a 25.0 ohm resistor are connected in series with an a. c. source whose e.m.f. is given by $E = 310 \sin 314\, t$ volt.

(*i*) What is the frequency of the e.m.f. ?
(*ii*) What is the reactance of the circuit ?
(*iii*) What is the impedance of the circuit ?
(*iv*) What is the current in the circuit ?
(*v*) What is the phase angle of the current by which it leads or lags the applied e.m.f. ?
(*vi*) What is the expression for the instantaneous value of current in the circuit ?
(*vii*) What are the effective voltages across the capacitor, the inductor, and the resistor ?
(*viii*) Construct a vector diagram for these voltages.
(*ix*) What value of inductance will make the impedance of circuit minimum ?

Ans : (*i*) 50 c/s
(*ii*) 96 ohm. net reactance is capacitive.
(*iii*) 99 2 ohm
(*iv*) 2.22 A
(*v*) The current leads the voltage by 75.4° or 1.31 radians.
(*vi*) 3.13 sin (314 t+1.31)
(*vii*) $V_0 = 282.6$ V, $V_L = 69.7$ V, $V_R = 55.5$V
(*ix*) 0.306 henry

14.20 Does the current in an a.c. circuit lag. lead or remain in phase with voltage of frequency f, applied to the circuit, when

(*i*) $f = f_r$
(*ii*) $f > f_r$
(*iii*) $f < f_r$

where f_r is the resonant frequency.

14.21 Prove that when a voltage $V = V_0 \cos 2\pi ft$ is applied across a resistor the average power dissipated per cycle in the resistor is given by

$$P = V^2_{eff}/R$$

where $V_{eff} = V_0/\sqrt{2}$

14.22 How much current is drawn by the primary of a transformer which steps down 220 volt to 22 volts to operate a device with an impedance of 220 ohm ? (0.01A)

14.23 Why can we not use the ordinary moving coil galvanometer for measuring a. c. current and voltage ?

UNIT 15

Universe

15.1 Introduction

The study of the universe is the object of astronomy which is perhaps the oldest of all sciences. The part of the universe in which we live is called the solar system. It includes the sun, the earth and the moon, other planets and their satellites, asteroids, comets and meteors. The outermost planet Pluto is at a distance of 6×10^{12} metres from us. Farther away are stars very much like our Sun, which occur singly, in pairs as binary stars, or in clusters like Pleiades (*Krittika*). We can see about 2500 stars at a time on a clear night, the total being about 5000 in the whole sky. They are called naked eye stars, but millions of stars can be seen through telescopes. The distances between stars are very large, the nearest star Alpha Centauri is at a distance of 4 light years. One light year is the distance travelled by light in one year with a speed of 3×10^8 m/sec, it is equal to 9.46×10^{15}m. The space between the stars is not completely empty; in it is spread very thinly interstellar matter consisting of gas and dust.

On a clear night we can easily identify a band of white light stretching across a great circle in the sky. It is called the Milky Way or *Akashaganga* The telescope resolves the band into millions of faint stars which form a large flat stellar system. The naked eye stars and our sun are parts of the Milky Way system containing more than a hundred billion (10^{11}) stars; hence we call it affectionately as 'our galaxy'. There are billions of galaxies like our own in the universe. Only the nearest ones can be seen with the naked eye, but photographs taken with large telescopes show millions of faint galaxies. The galaxies form the major building blocks of the universe.

Unlike other sciences, astronomy is an observational rather than an experimental science. In a way the 'experiments' are performed by nature on the surfaces and in the interiors of stars, and also in the interstellar space. We cannot control the conditions of these experiments as we can do in a laboratory. But this fact does not make the science of astronomy less interesting in any way. In fact the condition in the heavenly bodies are so varied and so extreme that we cannot create them in any of our laboratories on the surface of the earth. Therefore, it is very fascinating and re-assuring to discover that the laws of nature derived from our laboratory experiments are valid in all parts of the universe, in other words these laws are universal.

It is only in the last two decades that man

has been able to send space probes to our nearest neighbours in space viz., moon and the planets. But the only thing that we can observe about the stars and other heavenly bodies is the electromagnetic radiation (light, radio-waves, X-rays, γ-rays, etc.) coming from them. Thus we can find the direction from which the light is coming and changes in this direction; we can measure the amount of radiation and variation in it; we can note the colour and analyse the spectrum of light that comes to us; and we can measure its polarisation. From these we can derive a great many properties of the stars and planets such as : their brightness and distance, masses and radii, surface temperatures and compositions, conditions in the interiors, their motions and evolution. In this chapter our aim is to indicate some of these remote sensing methods employed and perfected by the astronomers.

15.2 Astronomical Instruments

Let us start by surveying the instruments used in astronomical research. They can be conveniently divided into two groups—instruments for collecting light or radio-waves are called telescopes; others are the recording instruments of various kinds. Our eye serves both as collector and recorder of light. But as collector it is not efficient on account of its small size and as recorder it is not reliable because of biological effects. Hence we need auxiliary equipment like telescopes and photographic plates.

Optical telescopes are of two kinds—refractors and reflectors (Fig. 15.1). In refractors a convex lens at the front end of the telescope collects the parallel rays coming from a distant object and brings them to focus at the focal plane of the lens. In reflectors the same purpose is served by a concave parabolic mirror. In either case the light collecting piece is called the objective. The light gathering power of the telescope, which determines our ability to see fainter objects, increases in proportion to the area of the objective $\frac{\pi D^2}{4}$, where

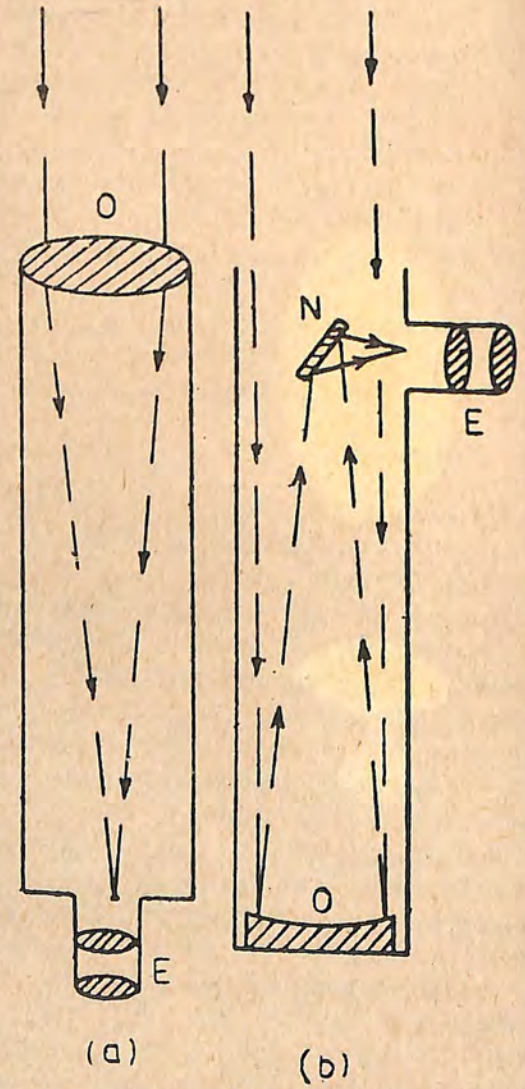


Fig. 15.1 Agtronomical telescopes
(a) Refractor type (b) Reflector type
O=Objective E=Eyepiece
N=Flat mirror

D is the diameter or aperture of the objective. Thus the 200-inch aperture telescope of the Palomar Observatory in U.S.A. collects 2500 times more light than a small telescope of 4-inch aperture. The details in the image, also increase directly in proportion to aperture and inversely in proportion to wavelength of light as explained in section 7.6. But in practice the fine quality of the image is finally limited by the turbulence and transparency of the atmospheric air. That is why astronomical observatories are normally located on remote mountain tops where the air is calm and clear.

Refractors suffer from chromatic aberration due to which the rays of all colours are not brought to focus at one point and the image becomes blurred. Reflectors, on the other hand, do not suffer from chromatic aberration and are easier to make. Hence the largest telescopes of the world, such as the 200-inch telescope of Mount Palomar in U.S.A. and 234-inch telescope of USSR, are all reflectors. The 48-inch telescope of Japal-Rangapur Observatory of Osmania University, which is the largest telescope in south-east Asia, is also a reflector. A 90-inch reflector is being made by the Indian Institute of Astrophysics, Bangalore for its Kavalur Observatory.

Radio telescopes, which are used for receiving the radio waves coming from heavenly bodies, are mostly reflectors of parabolic or cylindrical shape. On account of the larger wavelengths of radio-waves as compared to those of visible light, the radio telescopes require larger apertures for showing the same details as seen through an optical telescope. The parabolic disk of the large radio telescope in Jodrell Bank, U.K. is 250 feet in diameter. The objective of the big radio telescope in Ootacamund, India is a cylinder of 500 metres in length and 50 metres in width.

Photographic plate is most commonly used for recording the images of the sun, planets and star fields. These permanent records then serve for the measurement of positions and brightness of the stars, or for the study of the features on the disc of the sun or planet. Photographic plates have certain defects of their own; in particular the darkening of the plate is not proportional to the amount of light. Photocells are found to be very good for this purpose, because the photo-current is directly proportional to the intensity of light. But in their case we can observe only one star at a time, hence a judicial combination of photographic and photoelectric techniques is used in practice. Radio telescopes are normally coupled with special low noise radio receivers which operate at a fixed wavelength.

The infrared, ultraviolet, X-ray and γ-ray radiations of the heavenly bodies are absorbed by our earth's atmosphere. However, the advent of satellites and orbiting observatories has now made it possible to observe them from above earth's atmosphere opening up new branches of astronomy for the study of the universe.

15.3 The Study of the Solar System

It is known from the time of Kepler that planets move round the sun in elliptic orbits with sun at one focus of the ellipse. Motions of the planets under Newton's law of gravitation have already been considered in Unit 3. We shall now discuss how the physical properties of the objects in the solar system are obtained from observations. The results are given in Table 1.

(a) *Distance :* The method used by astronomers for obtaining the distances of heavenly bodies is the same as that used by a surveyor of land, viz., the method of triangulation mentioned in section 1.2. In the case of objects in the solar system the astronomer observes the object, say moon P, simultaneously from two widely separated places A and B on the surface of the earth (Fig 15.2). Noting the position of the object as seen from the two places he is able to measure the parallactic angle APB. Then knowing the positions of A and B with respect to earth's centre he can calculate the distance PO. For the moon the angle APB is found to be 57′ when AB is equal to the radius of the earth (6.378×10^6m). Hence putting $d=6.378\times10^6$ and $\theta=57'$ in equation (1.1) we get a distance of 3.844×10^8m for the moon.

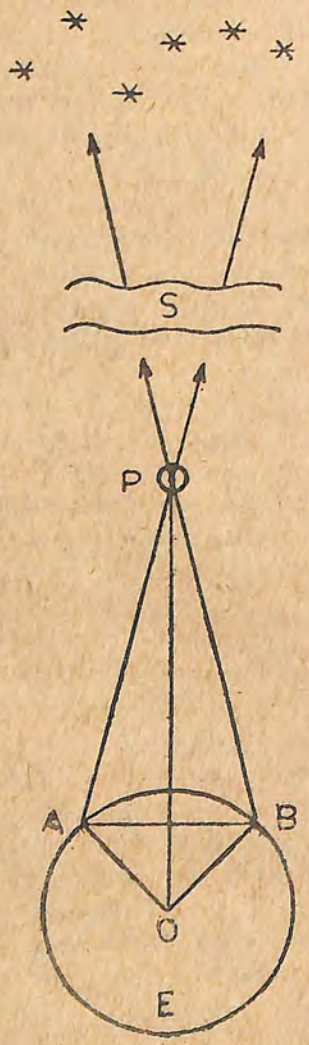


Fig. 15.2 Determination of distance of objects in solar system

For planets it is sufficient to make use of the above method for obtaining the distance of any one planet. The distances of all other planets then follow from Kepler's third law $a^3 \propto T^2$, where a is the semimajor axis and T the period of revolution. In recent years the distance of Venus has been accurately measured by the more direct method of radar echos. Radar signals sent to Venus travel with the speed of light 'c' and reach there after a time $t=r/c$, where r is the distance of Venus. On reflection from the surface of Venus they are received back at the radar station after an interval of time equal to 2t. The measurement of this interval thus gives $r=ct$.

Once the distance of Venus from the Sun at a given time is known, Kepler's third law can be used for deriving the distances of all planets from the Sun. It is thus that the mean distance of the earth from the sun, which is called the astronomical unit (A.U.), is found to be 1.496×10^{11}m. The distances of planets from the sun in A.U. are given in column 3 of Table 1 given at the end of the chapter.

(b) *Size :* From the definition of angle (Fig. 15.3) it follows that : D (size in m) = θ (angular size in radians) $\times$ r (distance in m) ... (15,1)

For example sun's angular radius is 960″ or 0.00465 radians. Multiplying it by the distance 1.496×10^{11}m we get a solar radius of 6.956×10^8m. It is about 109 times the radius

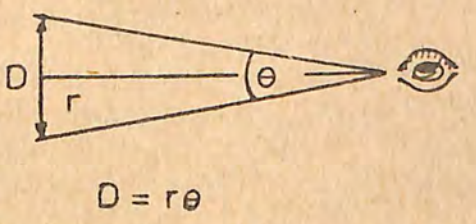


Fig. 15.3 Relation between linear size and angular size

of the earth. The radii of the planets in units of the radius of the earth found by this method are given in column 4 of Table 1. Similarly multiplying the angular separation of a satellite from the parent planet by the distance of the planet we can calculate the distance of the satellite from the planet in kilometres.

(*c*) *Rotation :* We know that our earth revolves round the sun in one year and at the same time it rotates around its north-south axis from west to east in 23 hours and 56.1 minutes. The planets also exhibit the phenomenon of rotation around an axis passing through their respective centres. The period of rotation of a planet is determined by observing the movement of surface markings or cloud patches over the disc of the planet. The rotation periods are given in column 5 of Table 1. In most cases the sense of rotation is from west to east as in the case of the earth. Mercury does not show any surface markings and Venus is always covered with a thick blanket of clouds. Their rotation periods were determined recently by using the radar echo technique. Both of these planets are found to rotate very slowly and the rotation of Venus is retrograde i.e., from east to west.

(*d*) *Mass :* Let M_1 be the mass of a large central body like the sun and let another smaller body, say a planet of mass M_2 revolve round it in a circular orbit of radius 'a'. Actually both bodies revolve round a common centre of mass C as shown in Fig 15.4. By the

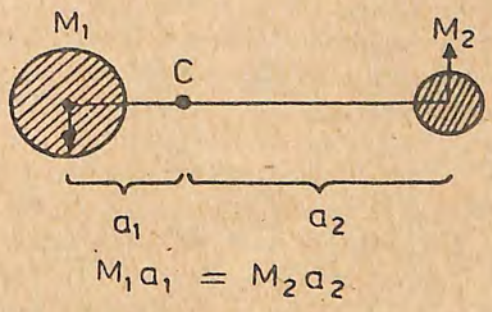


Fig. 15.4 Motion of a binary system around its centre of mass

principle of balance we have

$$M_1a_1 = M_2a_2, \text{ with } a_1 + a_2 = a$$

Then,

$$\frac{a_1}{a_2} = \frac{M_2}{M_1} \text{ which gives } \frac{a_1+a_2}{a_2} = \frac{M_1+M_2}{M_1}$$

Therefore,

$$a_2 = \frac{M_1\, a}{M_1+M_2} \text{ and } a_1 = \frac{M_2\, a}{M_1+M_2}$$

Now if V_1 and V_2 are the circular velocities of the two bodies in their orbits around the centre of mass we have $V_1 = \frac{2\pi a_1}{T}$ and $V_2 = \frac{2\pi a_2}{T}$ where T is the period. Then considering the force acting on one of the bodies, say M_2, we can write

Gravitational force = mass × centripetal acceleration

or,

$$G\frac{M_1M_2}{a^2} = \frac{M_2V_2^2}{a_2} = 4\pi^2\frac{M_2\, a_2}{T^2}$$

Dividing both sides by M_2 and substituting the value of a_2 in terms of a we find

$$\frac{GM_1}{a^2} = \frac{4\pi^2}{T^2}\,\frac{M_1\, a}{M_1+M_2}$$

It gives

$$M_1+M_2=\frac{4\pi^2}{G}\frac{a^3}{T^2} \quad \ldots (15.2)$$

Now in the case of the sun-planet system M_2 is much smaller than M_1, hence we can write equation (15.2) in the approximate form

$$M_1=\frac{4\pi^2}{G}\frac{a^3}{T^2} \quad \ldots (15.3)$$

Since the solar mass M_1 is a constant for all the planets, we see that $a^3 \propto T^2$, which is Kepler's third law of planetary motion. Mathematically it is found to be valid for elliptical orbits also if a is taken as the semi-major axis of the orbit. On applying equation (15.3) to the sun-earth system for which $a=1.496\times10^{11}$m and $T=365.25$ days $=3.156\times10^7$ secs we obtain, with $G=6.668\times10^{-11}$ MKS units.

M (sun) $= 1.989 \times 10^{30}$ kg
$=329000$ M (earth)

Equation (15.2) can also be applied to a planet-satellite system in which M_1 will be the mass of the planet and M_2 that of the satellite. Thus for Io, a satellite of Jupiter, $a=4.22\times10^8$m and $T=1\,769$ days $=1.525\times10^5$ sec; and equation (15.3) gives M (Jupiter) $=1.9\times10^{27}$ kg $=318$ M (Earth). This is the method of obtaining masses of planets, which have satellites. The masses of planets are given in column 6 of Table 1. On knowing the mass and radius we can easily calculate the mean density $=3M/4\pi R^3$ and acceleration due to gravity $g=GM/R^2$; they are given in columns 7 and 8 of Table 1.

EXAMPLE

For the moon, $M=7.349\times10^{22}$ kg and $R=1.738\times10^6$M; calculate its mean density and acceleration due to gravity at its surface. What will be the weight of a man on the moon if he weighs 60 kg on the earth ?

Solution

$$\text{Mean density of moon} = \frac{3M}{4\pi R^3}$$

$$= \frac{3\times7.349\times10^{22}}{4\times3.142\times(1.738\times10^6)^3}$$

$$=3.34\times10^3 \text{ kg/m}^3=3.34 \text{ g/cc}$$

$$\text{Acc. due to gravity on moon} = \frac{GM}{R^2}$$

$$\text{or } g\text{ (moon)} = \frac{6.668\times10^{-11}\times7.349\times10^{22}}{(1.738\times10^6)^2}$$

$$= 1.62 \text{ m/sec}^2 = \frac{1}{6}g \text{ (earth)}$$

The weight of a body $=mg$
Therefore,

$$\frac{\text{Weight on moon}}{\text{Weight on earth}} = \frac{g\text{ (moon)}}{g\text{(earth)}} = \frac{1}{6}$$

So weight of the man on moon $=60\times\frac{1}{6}=10$ kg.

(e) Surface temperature : Planets shine by the reflected light of the sun. Only a fraction of the solar radiation is reflected by the planet. The remainder is absorbed and it heats up the surface of the planet, which then radiates energy in the infrared and radio regions of the electromagnetic spectrum. On measuring the total self radiation of the planet and applying Stefan's law $E=\sigma T^4$ we can determine the surface temperature T of the planet. Column 9 of the Table 1 gives the temperatures of the planets obtained in this way. The observed temperatures decrease as we go away from the sun, because the planets receive less and less solar energy according to inverse square law. Venus is an exception, because it has a very thick atmosphere of carbon-di-oxide which acts as a blanket and keeps the surface hot.

(f) Atmosphere : We know that our earth has an atmosphere consisting of 80% nitrogen and 20% oxygen with some carbon-di-oxide and water vapour. We can determine whether a planet or satellite has an atmosphere in two ways. One is to observe meteorological phenomena of clouds and storms; the other is to detect atmospheric gases with the help of a spectrograph. As clouds are good reflectors

of light they increase the reflecting power of the planet, which is called albedo. Column 10 of Table I gives the albedos of various planets and the moon. We see that the Moon and Mercury reflect only 6 to 7 per cent of sunlight very much like the rocks on earth. Hence we can conclude that these bodies have no atmosphere which is indeed verified by space probes. The planets Venus, Earth, Jupiter, Saturn, Uranus and Neptune have high albedos, which indicate that they have cloudy atmospheres. Venus has the highest reflecting power of 85% because it is always covered by thick clouds all over its surface. The atmosphere of Venus is so thick that it produces a surface pressure 100 times that on earth. Mars and Pluto have intermediate albedos which indicate that they have thin atmospheres. In the case of Mars the atmospheric pressure is only six-thousandth of our earth's atmospheric pressure.

The chemical composition of the atmosphere of a planet is derived from the study of its spectrum which would show the characteristic absorption bands of the gaseous molecules in the atmosphere, (refer to section 7.9). This is how it was found that the atmospheres of Venus and Mars contain mostly carbon-di-oxide. Atmospheres of Jupiter and Saturn contain hydrogen, helium, methane and ammonia, while those of Uranus and Neptune have hydrogen, helium and methane only.

We may wonder as to why some planets have atmospheres of varying thickness while others have no atmosphere at all. There are two factors, the gravity (g) and temperature, which come into play. Small bodies like the moon have very low gravity and consequently small escape velocity (see section 3.11). Therefore the atoms of atmospheric gases escape one by one leaving nothing behind. Mercury, Mars and Pluto, for which g is somewhat larger, should retain some atmosphere. Only in the case of Mercury which is very hot on account of its closeness to Sun, the atmosphere has completely escaped. As discussed in section 8.5, high temperature is associated with larger kinetic energy and hence larger mean velocities of atoms and molecules, which are sufficient to overcome the effect of gravity. Again since the velocity of lighter atoms is more by the law of equipartition of energy discussed in section 8.6, they can escape much more easily than the heavier ones. That is why the planets, Venus, Earth and Mars have lost all of their hydrogen and helium which are the lightest of all atoms. Heavier planets Jupiter and Saturn have, however, retained all gases including hydrogen and helium.

15.4 Space Exploration and Search for Extra-Terrestrial Life

Space exploration started with the launching of the first sputnik by Russia on 6 October 1957. Manned space flights began with the circling of the earth by Gagarin on 12 April 1961. The aim of all the space effort has been to gain knowledge about the objects in the solar system at first hand and to search for life on them. In the beginning earth satellites were launched for studying the upper atmosphere of the earth. One very important discovery made by the satellite Explorer I in January 1958 was that of Van Allen radiation belts of charged particles trapped in earth's magnetic field. In addition much information about the density, temperature and composition of earth's atmosphere upto a height of 1000 km was obtained. Thereafter the space scientist ventured to step out towards the moon and the planets.

(*a*) *Moon* : The moon is the only natural satellite of our earth. As it is our closest neighbour in space it has been extensively studied by space probes. The moon always keeps the same face towards the earth because

its period of rotation is equal to its period of revolution round the earth. The rear side of the moon was first photographed by the Russian space probe Luna-3 in 1960. It also shows craters, dark planes and lighter shaded mountains like the visible front side. From telescopic studies it was already known that the lunar landscape is bleak and foreboding. It is an airless and waterless expanse of desert and rocks without vegetation, life and sound. Lack of atmosphere, and long duration of the day, which lasts for 15 earth days there, makes the temperature soar to 110°C at mid-day and drop to −150°C at night. Hence the U.S.A. first sent out surveyor and orbiter space crafts to observe the moon from close range before the astronauts Armstrong, Aldrin and Collins finally conquered it in July 1969 in their Apollo 11 space vehicle. They and their successors brought back lunar soil and rocks for examination in our laboratories. The Russians also sent an unmanned robot vehicle Lunakhod in January 1973 and brought back lunar samples to earth all by remote control. Analysis of lunar material has shown that it is similar to the rocks in earth's crust. The age of lunar rocks agrees with the age of the earth and meteolites, all being about $4\frac{1}{2}$ billion year old. We can conclude that all objects in the solar system were formed more or less simultaneously.

(*b*) *Mercury* : Mariner 10 space vehicle of U.S.A. passed within a few thousand kilometres of Mercury in February 1973. Photographs taken at that time show that Mercury is pock-marked by craters of all sizes just like the moon. No life can exist on Mercury due to lack of any atmosphere and extremes of temperature.

(*c*) *Venus* : In radius, mass and density Venus is almost a twin of earth. So the thick cloud cover on its surface suggested that it could have water and life similar to earth. But the Venera space probes sent by Russians between 1965 and 1971 showed that Venus is covered with a very thick atmosphere containing 95% carbon-di-oxide The surface pressure is 100 times the terrestrial atmospheric pressure and the surface temperature is 480°C. There is some water, 0.7 per cent, in the atmosphere of Venus ; but the material of clouds is quite different. Under these circumstances there is no possibility of life on Venus.

(*d*) *Mars* : Mars with its red soil-like colour, nearly twenty four hour rotation period, yellow clouds, blue hazes and white polar caps, appears so much earth-like that it was considered a fit abode for man-like beings. Changing colour with the march of seasons on Mars has so often been mistaken as a sign of vegetation. So USSR sent Mars vehicles and U.S.A. sent Mariner space probes to Mars from 1971 onwards for the study of this fascinating planet. Spectroscopists had already concluded that the thin atmosphere of Mars is mostly made of carbon-di-oxide. They had also found traces of water and believed that the polar caps might be made of water ice. However, observations made from space crafts showed that the caps were made of dry ice, i.e., solid CO_2 with only a few per cent of water in them. In spite of these dry, cold and oxygenless conditions it was thought that primitive forms of life, such as lichens, might occur on Mars.

Close range photographs taken by Mariner 9 during 1971-72 showed that Mars has craters like the moon. But the boundaries of craters are smoothed by the eroding action of wind which raises huge dust storms and produces sand dunes on Mars. Further it was found that Mars shows geological activity such as volcanic eruptions, formation of huge canyons and raised plateaus. Photographs also revealed dry river beds which indicate that sometime in the past Mars was warmer and water flowed

freely over its surface. At present Mars appears to be going through a cold glacial period with all water frozen as permafrost under the layer of surface rocks. Volcanic activity could bring this water to surface at some low lying warm places where life could thrive. The Viking 1 and 2 space crafts were sent by U.S.A. to find out whether there is any form of life on Mars. The first Viking soft landed on Mars on 20 July 1976. It took excellent pictures and conducted three experiments on the martian soil to detect life. Unfortunately the results have been rather negative. It appears that our earth is perhaps the only abode of life in the solar system. Of course there is a good chance that life might exist on planets moving around some distant stars.

(*e*) *Outer planets* : All planets beyond Mars are so cold that living organisms are not likely to survive there. Further their atmospheres are full of poisonous gases like methane and ammonia. Anyway Pioneer and Voyager space crafts have been launched by America for a close study of the outer planets with the hope that we could at least understand the origin of our solar system.

(*f*) *Other objects in the solar system* : Besides the planets and their satellites the solar system contains three other types of bodies viz., asteroids, meteors and comets. Asteroids are a group of more than 1600 small bodies moving around the sun between the orbits of Mars and Jupiter. The largest among them is Ceres with a radius of 350 km while the smallest might be only 100 metres across. They are believed to be the pieces of a much larger planet which broke up due to the gravitational effect of Jupiter. The density and chemical composition of asteroids are very much like that of the moon.

Comets are loose balls of ice made of water, ammonia and methane with solid metallic particles embedded in them. They move in highly elliptical orbits and spend most of their time far away from the sun. A comet has a nucleus of about 10 km diameter which is not visible most of the time. But when it approaches the sun solar radiation heats up the material, which evaporates and forms a head of about 10,000 km diameter. On still closer approach to sun the comet developes a tail pointing away from the sun due to pressure of sunlight. Many comets are observed to break up into smaller pieces when they approach very close to the sun. When earth's orbit crosses the orbit of the comet, this material falls on earth in the form of a meteor shower. Most of the meteors are small and they burn up due to friction in earth's atmosphere. But some larger pieces survive and fall on the ground as stony or iron meteorites. The biggest of them form craters on the surface of the earth. The cause of craters on Moon, Mercury and Mars is also the same.

Space missions to asteroids and comets are being planned by America during 1980's.

15.5 Study of the Sun

Our sun is a typical star of average size, mass and brightness. It looks brightest because it is only 8 light minutes from earth. This enables us to study it in great detail compared to other stars which are so far away that they appear as point sources of light even through the largest telescopes. We have already seen in section 15.3 how the distance and mass of the sun are determined. We shall now consider how we can find its luminosity, temperature and other properties.

(*a*) *Surface temperature* : We can measure the amount of radiation coming from the sun by noting the rise in the temperature of a black surface exposed to the sun (Fig. 15.5). It is found that we receive on earth every minute

about 2 calories of solar radiation on each square centimetre area held perpendicular to

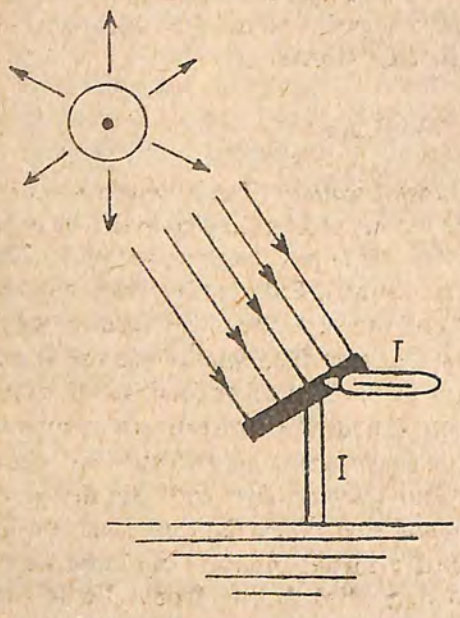


Fig. 15.5 Measurement of solar constant
T—Thermometer

the rays of the sun. This is known as solar constant, it amounts to 1.388×10^3 watts/m^2.

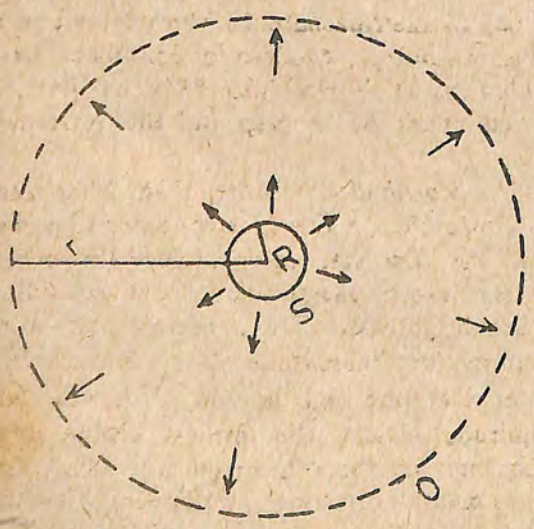


Fig. 15.6 Calculation of solar luminosity and temperature

Now consider a sphere of radius r equal to one astronomical unit, which is the distance of the earth from the sun (Fig. 15.6). The total radiation crossing its surface area ($4\pi r^2$) comes out to be 3.9×10^{26} watts ; this is the total luminosity of the sun. Again dividing it by $4\pi R^2$, where R is the radius of the sun, we find that the flux E emitted by one square metre of solar surface is equal to 6.41×10^7 watts/m^2. Then using the Stefan-Boltzmann law $E = \sigma T^4$. where $\sigma = 5.669 \times 10^{-8}$ MKS units, we find that the surface temperature of the sun is

$$T = \sqrt[4]{\frac{6.41 \times 10^7}{5.699 \times 10^{-8}}} = 5800 \text{ Kelvin}$$
$$= 5527°C$$

No material can exist in solid or liquid form at this temperature.

(b) Conditions at the centre : The temperature, pressure and density increases as we go deeper below the surface layers of the sun on account of the compression by the weight of the overlying material. Calculations show that the temperature at the centre of the sun is about 14 million (14×10^6) Kelvin. It is obvious that the whole body of the sun is gaseous in form. The gas is mostly hydrogen, which accounts for 70 per cent of the mass of the sun. The remainder is made up of 28 per cent helium and 2 per cent all other heavier elements from lithium to uranium. The density of the sun also increases from about 10^{-4} kg/m^3 at the surface to more than 10^4kg/m^3 at the centre. The mean density calculated from its mass and radius is 1.4 times the density of water. The pressure at the centre of the Sun is 2×10^{16} N/m^2.

(c) Energy production : Dividing the luminosity of the sun by its mass we find that every kg of its material emits nearly 2×10^{-4} J of energy every second. During the total life of the sun, which is $4\frac{1}{2}$ billion years old, each

gram has emitted 3×10^{10} J of energy. Such tremendous amount of energy cannot be produced by chemical reactions such as burning of coal and oil or by explosion of gunpowder. In 1939 Hans Bethe showed that controlled thermonuclear reactions are responsible for the generation of solar energy. Two kinds of reactions have been suggested. They are known as carbon-nitrogen cycle and proton-proton chain reaction ; the latter reaction occurs in the hydrogen bomb. In both of them 4 hydrogen atoms fuse together to form a helium atom. Now the atomic weight of 4 hydrogen atoms is 4.032 while that of one helium atom is 4.003. Thus there is a loss of 0.7 per cent of mass and according to Einstein's equation $E=mc^2$ it gives rise to 5×10^{14} J of energy per kg which is more than sufficient to keep the sun shining for a very long time.

(*d*) *Sunspots and solar activity* : Photographs of the sun taken in white light show many dark spots which are called sunspots. Sunspots appear dark by contrast with the solar surface, because they have a somewhat lower temperatures of about 4500K. Some of the sunspots are so big that they can easily swallow several earthlike bodies. Sunspots are characterised by intense magnetic fields of 2 to 3 thousand gauss. The number of sunspots are found to vary from year to year with a period of about 11.1 years. This periodicity is known as sunspot cycle. The sunspots are found to move across the disc from day to day; from this it is inferred that the sun rotates with a period of about 25 days at its equator. Quite often there is a sudden increase of intensity in the red H_{α} light at some places on the solar disc, this phenomenon is called a solar flare. During a flare the sun emits streams of protons, electrons and alpha particles which reach the earth about a day later causiug global magnetic storms and radio disturbances. All this solar activity also varies with the period of sunspot cycle. It is found that the growth of trees is also affected by the sunspot cycle, which is seen as a variation in the thickness of the rings in their trunks.

15.6 The Stars

(*a*) *Identification* : The brightest stars in the sky are recognised by their names such as Sirius (*Vyadha*), Canopas (*Agasti*), Spica (*Chitra*) Arcturus (Swati), Polaris (*Dhruva*), etc. Many bright and faint stars together form groups called constellations. Examples of constellations are the signs of the Zodiac or Rashis such as Aeries (*Mesha*), Taurus (*Vrishabha*), etc., which lie on the apparent path of the sun in the sky. Great Bear (*Saptarishi*) and Southern Cross (*Trishanku*) are constellations near the north and south celestial poles. In India we have been using *Nakshatra* from Vedic times. They are also located close to the apparent paths of the sun and the moon in the sky. Since the moon takes 27.32 days to make one round, among the stars there are 27 *Nakshatras* one for each day. That is the basis of naming the day by the *Nakshatra* in which the moon is found at sunrise. Modern astronomers have divided the whole sky into 88 constellations like countries on a map of the terrestrial globe.

(*b*) *Magnitudes* : More than 2000 years ago Hipparchus, a Greek astronomer who prepared the first star map, divided the naked eye stars into six magnitude classes according to their brightness. The brightest stars were put in the first magnitude class. Successively fainter stars were put in 2nd, 3rd, 4th, 5th magnitude classes, the faintest visible stars being put in the 6th magnitude class. On precise measurement with photometers attached to telescopes it was found in the 19th century that the ratio of brightness between successive magnitude classes is 2.512, so that a

1st magnitude star is $(2.512)^5 = 10$ times brighter than a 6th magnitude star. On using this relation it was found that some of the stars originally put in the 1st magnitude class are actually brighter than the average for that class. Hence zero and negative magnitudes were introduced. On the other hand more powerful telescopes made fainter stars visible. Therefore, it was necessary to extend the magnitude scale beyond the 6th. Thus the stars 2.512 times fainter than the 6th magnitude were assigned 7th magnitude and so on. The 200 inch telescope can show us stars of 22nd or 23rd magnitude. It is expected that the 90 inch space telescope to be put in orbit around the earth by the space shuttle in 1980's would be able to show us stars of 28th magnitude, which would be 10^{22} times fainter than our sun.

(c) *Distance and absolute brightness of stars*: So far we have been talking of the brightness of a star as it appears to us or in other words its apparant magnitude. A star looks bright either because it is intrinsically bright or because it is nearer to us. The sun looks brightest of all the stars because it is closest to us, although it is really fainter than many other stars like Batelgeuse which look fainter because of their large distance. The true or absolule magnitude of a star corresponds to the brightness of the star if it were taken to some standard distance.

The distances of stars are also measured by the method of triangulation described in section 1.2. Distance of some of the celestial objects are given below :

Object	*Distance*
Alpha Centauri	4.35 light years
Sirius	8.48 ,, ,,
Pleiades (*Krittika*)	410 ,, ,,
Globular cluster M13	27000 ,, ,,
Andromeda galaxy	2 million light years

The apparent brightness of an object varies inversely as the square of the distance. So we can compute the true brightness of a star from its observed apparent brightness and distance by bringing it to a standard distance.

On this basis we find that Sirius and Betelgeuse are respectively 22 and 19000 times brighter than our sun intrinsically. They look faint only on account of their large distances. The absolute brightness of stars ranges between 60000 times to (1/400000)th of the brightness of the sun.

15.7 Physical Properties of Stars

(a) Stellar spectra: The spectrum of the sun has been described in section 7.9. Just like the sun, the stars also show a continuous spectrum on which dark absorption lines are superimposed. The continuous spectrum is produced by the surface layers of the star which form what is called the photosphere of the star. The photosphere emits continuous radiation in all wavelengths like a black body. The dark absorption lines are produced when the continuous radiation of the photosphere

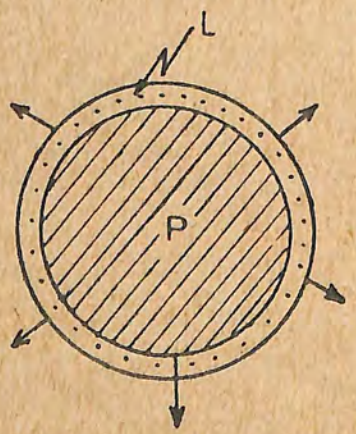


Fig. 15.7 Emission of continuous spectrum and formation of absorption lines in stellar spectra

passes through the top most cooler layer which is known as the reversing layer (Fig. 15.7). The wavelengths of the dark absorption lines

depend upon the composition of the material of the reversing layer. So we can identify the atoms and molecules in the stars by noting the positions of Fraunhoffer lines in their spectra. It is found that stars contain the same 92 natural and about a dozen artificial elements that are found on earth.

The spectra of all stars are not alike; the spectrum is related to the colour of the star. There are seven principle classes of stellar spectra denoted by letters O, B, A, F, G, K and M. Their properties are given in Table 2. We see that O and B type stars are blue and their spectra show helium lines. The A type stars are white and their spectra are dominated by hydrogen lines. F and G type stars of green and yellow colours show lines of ionised and neutral metals. Finally the spectra of the orange K and the red M stars contain molecular bands of titanium-oxide in their spectra.

(b) Interpretation of stellar spectra : The colour of a star is actually related to the temperature of the photosphere This can be understood from Wien's displacement law: λ (max) T=constant. As the temperature is increased λ(max), the wavelength at which the intensity is maximum, becomes shorter and the colour changes accordingly. For example, if we heat an iron ball, which is black when cold, it first becomes red hot and on further heating changes its colour to white and then to blue. Similarly the coolest stars look red or orange, the hotter ones appear yellow or green while the hottest stars are white or blue. Thus we see that the spectral sequence O, B, A, F, G, K,M arranges the stars according to decreasing surface temperature.

In earlier days it was thought that the differences in the spectra of stars were caused by the differences in their chemical composition. For example O and B stars were called helium stars, A stars were thought to be made of hydrogen and F and G stars of metals. The Indian astrophysicist M. N. Saha showed in 1922 that all stars have the same chemical composition and the differences in their spectra are principally due to different surface temperatures of stars. At low temperatures of a few thousand degrees, the gases occur in the form of molecules as is found to be the case for the M type stars. As we increase the temperature the molecules first break into neutral atoms and on further increasing the temperature the atoms become ionised. Metals have low ionisation potentials, hence they are ionised at lower temperatures. This is what happens as we go from M and K to G and F type stars. By the time the temperature is raised to about 10000 degrees in A type stars all metals lose several valence electrons and hydrogen is excited giving rise to strong Balmer lines. By further increase of temperature to 20000 degrees, hydrogen is completely ionised and helium is excited. Thus in B stars neutral helium lines dominate the spectrum. Finally in the hottest O type stars with temperatures of 35000 degrees even helium is ionised. Thus Saha's theory of ionisation has helped us to determine the surface temperatures of stars from their spectra as given in column (5) of Table 2.

After knowing the temperature it is easy to find the true chemical composition of the stellar atmospheres. H. N. Russell was the pioneer in this field; he and later workers found that all stars have essentially the same composition as that of the sun. Not only stars but the gas and dust that exist in the interstellar space are also made of the material of the same composition. This cosmic mixture has hydrogen, the lightest of all elements, as the most abundant constituent. 88 per cent of all the atoms in the universe are hydrogen atoms, 11.8 per cent are helium atoms and the remaining heavier elements make up only

0.2 per cent population of atoms.

(*c*) *Radii of stars* : Since the stars are seen as point sources even with the largest telescope their angular diameters cannot be measured. So the method of determining size described in Section 15.3 cannot be applied to them and we have to use some other indirect method. If we assume that stars emit radiation like black bodies then each unit area of its surface would emit a Flux $F=\sigma T^4$ where T is the surface temperature. The surface area of the star is $4\pi R^2$ where R is the radius of the star. Hence the total luminosity of the star will be $L=(4\pi R^2)\times(\sigma T^4)$. This equation will give the radius R when temperature T is known from the spectrum and luminosity L from the absolute magnitude. The radii of various spectral types of stars are also given in column (8) of Table 2. For most stars they range from about 20 solar radii to about 1/10th of the radius of the sun. There are a few exceptions : for example Capella and Betelgeuse have radii equal to 50 and 220 times the radius of the sun. Such stars are called giants and supergiants on account of their large size. By contrast the other most common stars like the sun are called dwarfs. At the other extreme we have stars like the companion of Sirius which is only twice as big as the earth in diameter. These very small stars are called white dwarfs.

(*e*) *Masses of stars*: About half of the stars in the milky way are binary systems. A binary system contains a pair of stars which revolve around their common centre of mass. α Centauri, Sirius, Spica are all binaries. Masses of stars are obtained by the study of binaries and through the application of Kepler's third law incorporated in equation (15.2). For convenience let us take sun's mass as the unit of mass, the astronomical unit as the unit of distance and the year as the unit of time. In these units $G=4\pi^2$ and equation (15.2) can be written simply as

$$M_1+M_2=\frac{a^3}{T^2} \qquad \ldots (15.4)$$

Let us consider Sirius as typical example of a binary system. It has two components Sirius A and Sirius B which are separated by a mean distance $a=7''.6=3.7\times10^{-5}$ radians. Since the distance of the binary is 8.5 light years $=5.4\times10^5$ A.U. an application of equation (15.1) gives

$$a=3.7\times10^{-5}\times5.4\times10^5 \text{ A.U.}=20.3 \text{ A.U.}$$

From observations it is found that the two components revolve around their common centre of mass in 49.9 years. Then substitution in equation (15.4) gives

$$M_1+M_2=\frac{(20.3)^3}{(49.9)^2}=3.35 \text{ M (sun)}$$

Further by locating the centre of mass, it is found that $a_2/a_1=2.35$. Hence we get $M_1=2.35$M (sun) and $M_2=1.00$M (sun). Masses of about 100 stars have been measured by the above method. They are found to be correlated with spectral type as shown in column (7) of Table 2. The masses of stars range from about 40 times the mass of the sun for the hottest O stars to a tenth of the solar mass for the coolest M stars.

For all stars except the white dwarfs it is found that the luminosity increases with mass roughly according to the $3\frac{1}{2}$th power of mass. For example Sirius, which is 2.3 times as massive as the sun, is 25 times brighter than the sun. The heaviest stars of 40 solar masses are a million times brighter than the sun while the least massive stars of one tenth solar mass are 10000 times fainter than the sun. (See columns (9) and (10) of Table 2).

(*f*) *Interiors of stars* : On dividing the mass of a star by its volume we get its mean

density. The mean densities of dwarfs stars range from 10 Kg/m³ for the hottest O stars to 50000 Kg/m³ for the coolest M stars. The lowest densities are found in supergiants e.g. Antares (α Scorpii) is only three thousandth as dense as air. The densest stars are white dwarfs like the companion of Sirius which is 5000 time denser than the densest element platinum. It is often said of these stars that one matchbox full of their material will weigh one ton on the surface of the earth.

As in the case of the sun, temperature and density increase towards the centres of stars. The central densities of dwarf stars range from a few thousand Kg/m³ for O type stars to 10^5 Kg/m³ for F,G,K and M type stars. The central temperature is maximum (about 30 million degrees) for early type O and B stars and minimum (about 10 million degrees) for M type stars.

15.8 Stellar Evolution

Like every thing else in the universe stars are born live for a certain length of time and die at the end of their career. The life history of a star begins when a large cloud of interstellar dust and gas begins to contract under the influence of its own gravitational force. It is calculated that the cloud must have a mass of at least one thousand solar masses for this to occur. As the cloud contracts it heats up by compression A part of the energy is radiated away, which helps the cloud to contract further. At some stage the cloud breaks up into a large number of fragments of stellar size and each fragment continues to contract. When fragments become hot enough, they begin to radiate light from their surface. Thus each fragment becomes a self-luminous star and a cluster of stars is born.

Each star of the cluster continues to contract until the central temperature attains value of ten million degrees or more. At these temperatures the thermonuclear reactions discussed in section 15.5 start. The carbon-nitrogen cycle and the proton-proton chain reaction convert hydrogen into helium and the energy liberated in these processes keeps the star shining for millions of years. The dwarf stars, which form more than 90% of the whole stellar population, are found to be in this stage of evolution. It lasts until all hydrogen fuel in the central ten per cent core of the star is exhausted The more massive stars of O and B type spend their fuel most rapidly like a spendthrift within 10 to 100 million years. The least massive M type stars, which are more frugal can remain shining for more than thousand billion years. The sun has a total life of 10 billion years out of which about half is already over at the present time.

When the hydrogen in the core of the star is exhausted the core begins to contract while the outer regions expand. In this way the radius of the star increases while the surface temperature drops and the star becomes giant or supergiant. For the sun this stage will occur about 5 billion years from now.

The time spent by a star in the giant and supergiant stages is small compared to the duration of the dwarf phase. That is why we find most stars in the dwarf stage and only a few in the evolved stages or giants and supergiants. At the end of the latter stage the production of energy in the star is so large and rapid that the star explodes in the form of a nova or supernova throwing out a large portion of its envelope into interstellar space. The core that remains behind may end up as one of the following three types of stellar remnants or corpses :

(i) If the original mass of the star was less than about 2 solar masses we get a dense white dwarf of less than 1.2 solar mass. As there is no nuclear fuel left in the white

dwarf it just cools off slowly changing its colour from white to yellow, to red and finally becomes a dark body.

(ii) If the original mass of the star was between 2 and 5 solar masses, the back kick of the supernova explosion will compress the core of the star to nuclear densities giving rise to a neutron star. The mass of a neutron star is less than 2 solar masses and its radius is about 10 kilometres. Neutron stars have very large magnetic fields of the order of 10^{12} gauss. If the magnetic axis is inclined to the axis of rotation, the star emits pulses at regular intervals, the periods of which range from 30 milliseconds to 3 seconds. These are pulsars. the first of which was discovered by the radio astronomers in 1967.

(iii) If the original mass of the star was more than 5 solar masses, the back kick of the supernova explosion is so violent that the core continues to contract indefinitely giving rise to a black hole. As the contraction proceeds, the radius decreases continuously and acceleration due to gravity g, at the surface goes on increasing. Finally a stage comes when the g value is so large that even the photons cannot escape from the surface of the body. On the other hand any particle or photon approaching it will be immediately swallowed. That is why such a body is called a black hole. The recently discovered X-ray source Cygnus-XI is found to be a binary system in which one component is believed to be a black hole

15.9 The Milky Way

(a) *Size and shape* : We nave noted earlier that our solar system and most of the naked eye stars are members of a very large system which appear to us as Milky Way (*Akashaganga*) in the sky. It is a flat lens shaped disc which is thicker near the centre and thins out towards the edges. We are situated very nearly in the central plane of the milky way. That is why it appears to us as a great circle in the sky. But the milky way is not uniformly bright everywhere. It is brighter and broader in one hemisphere, and less conspicuous in the other. This is due to the fact that the sun is not at the centre of the galaxy, the centre lies in the direction

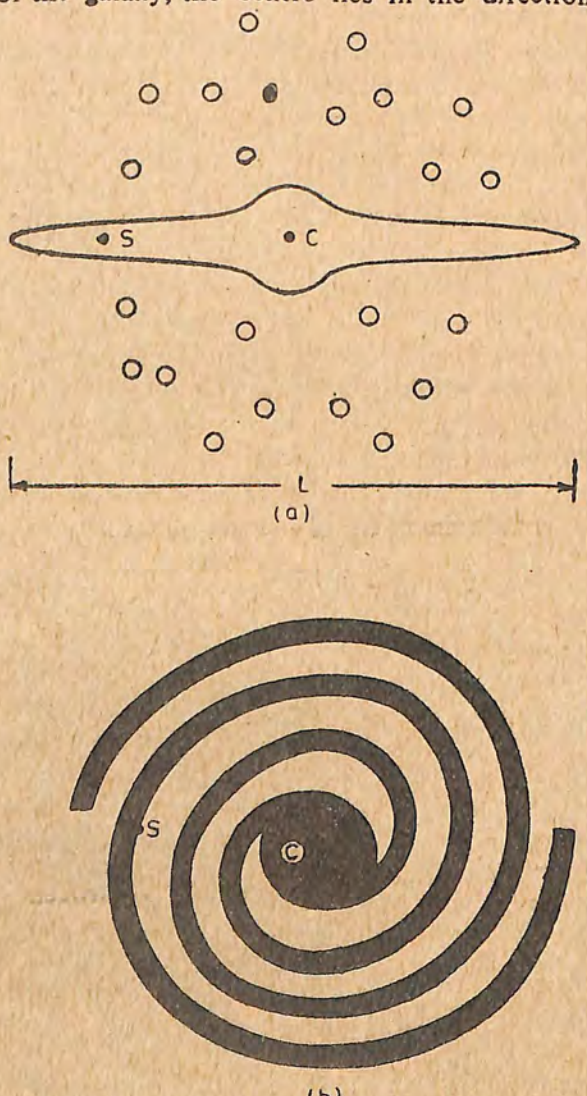


Fig. 15.8 (a) Schematic view of Milky Way i.e. Edge on view C=galactic centre S=position of Sun
(b) Schematic view of Milky Way i.e. Face on view C=galactic centre E=position of Sun

of the Sagittarius (*Dhanur Rashi*) constel lation. The diameter of the galactic disc is 100000 light years and the sun is situated at a distance of 30000 light years from the centre. The thickness of the disc is 5000 light years at the centre, about 1000 light years near the sun, and diminishes as we reach the edges (Fig. 15.8).

(*b*) *Interstellar matter* : On a clear night we see several dark regions in the band of the milky way. These regions look dark not because there are fewer stars in those directions, but on account of the intervening dust clouds which hide from our view the stars lying behind them. Dust and gas occupy the spaces between the stars and sometimes form dense clouds containing 10^8 to 10^9 atoms per m^3. But these densities are minute compared to the density of our earth's atmosphere. Like fog on the surface of the earth the dust in the interstellar space diminishes the light of the stars behind them And just as the sun looks red at the time of rising and setting because its light has to pass through a great thickness of our atmosphere, similarly the light of the star becomes reddened as it passes through the dust clouds which lie in its path. Sometimes a dust cloud is illuminated by a hot star and it begins to shine by the reflected light giving rise to a bright nebula like Orion nebula. The gas which forms 90 per cent of the mass of the interstellar matter can be recognised by its characteristic spectrum lines which appear either in emission or in absorption. Hydrogen is the most abundant of all the elements and it gives rise to an emission line in the radio region of the spectrum; it is called the 21 cm line which is very useful for locating the hydrogen clouds in our galaxy.

(*c*) *Star clusters* : The galaxy also contains smaller assemblies of stars known as clusters. Two kinds of star clusters are recognised. The open clusters, of which the Pleiades (*Krittika*), Hydes (*Rohinishakata*), Praescape (*Pushya*) are good examples, contain 100 to 1000 stars which appear separate on photographs. Quite distinct from galactic clusters are globular clusters which derive their name from their spherical appearance. A globular cluster contains about 100000 stars which are packed rather closely together. About 100 globular clusters are known, all of them are more than 20000 light years away from the sun.

(*d*) *Structure* : It is now known that the disc of our milky way has a spiral structure like what we see in the Andromeda galaxy (Fig. 1.1). Ordinary optical telescopes are inadequate for the study of its detailed structure. Our view is cut off by the interestellar dust clouds so that only a small portion in the solar neighbourhood can be surveyed. But radiowaves can penetrate through the dust clouds; therefore a detailed study of the galactic structure has been made in recent times by observations of the 21 cm line of hydrogen with radio telescope.

The central part of our galaxy looks like an amorphous spheriod and it resembles a huge globular cluster in many ways. Outside this central nucleus are situated the spiral arms which are about 1200 light years wide and are separated by gaps of about 500 light years. The sun is situated near the inner edge of one such arm which passes through the Orion nebula. The spiral arms contain the youngest objects in the milky way. They include the gas and dust, the open clusters and the hot blue stars. The inter-arm regions contain somewhat older stars. But the oldest stars are found in the galactic halo and in globular clusters. Fig. 15.8 shows schematically two views of our galaxy, one edge on and the other face on.

(*e*) *Rotation and mass* : The milky way is not stationary, it is rotating around an axis

passing through its centre. This is the real reason for its flat disc like shape. The globular clusters do not take full part in rotation, hence they form a spherical halo around the galaxy. The galaxy does not rotate like a solid wheel. Each star revolves around the central nucleus in an elliptical or circular orbit just like the planets round the sun. The stars near the centre move faster than the stars farther out. Our sun moves round the centre of the galaxy with a speed of 250 km/sec. The sun takes about 240 million years to comp ete one revolution around the galactic nucleus, which can be called the galactic year. Here also, equating the centrifugal force caused by the sun's circular velocity with the gravitational force of attraction by the central nucleus we can write : M (Galaxy) $a = v^2/G$. Putting $a = 30{,}000$ light years $= 3 \times 10^{20}$m, $v = 2.5 \times 10^5$m/sec. and $G = 6.67 \times 10^{-11}$ MKS units, we find that the mass of the galaxy is 3×10^{41}kg $= 150 \times 10^9$ solar masses. Hence, if we assume that the galaxy is made of average stars like the sun we can conclude that there are 150 billion stars in our galaxy.

15.10 Galaxies and the Universe

(*a*) *Normal galaxies*. Stellar systems like our Milky Way form the major building blocks of the universe; they are called galaxies. Hundreds of millions of galaxies have been photographed by the modern powerful telescopes. The farthest of them are about a billion light years away.

All galaxies do not look alike, and on this basis they are divided into three classes. About 3 per cent of all galaxies are irregular in shape, they are called irregular galaxies. A large number of galaxies show spiral arms, they are known as spiral galaxies. Our Milky Way and the Andromeda galaxy are good examples. Most common are elliptical galaxies which show elliptical discs on photographs.

(*b*) *Radio galaxies and quasars* : Normal galaxies emit a moderate amount of radio radiation compared to their light output. But certain radio sources in the sky are found to emit millions of times more energy in the radio region compared to normal galaxies. They can be divided into two types : radio galaxy and quasars. Radio galaxies are identified with some peculiar optical galaxies showing explosions in them. But the radio radiation does not come from the galaxy itself. Invariably we find two radio sources occuring symmetrically on either side of the peculiar galaxy, like two ears on the two sides of the face of a man (Figure 15.9). It is

Fig. 15.9 A=Radio galaxy G=Central optical galaxy S_1, S_2=Two parts of radio source

believed that a great explosion occured in the central galaxy and two clouds of charged particles were ejected on two dimetrically opposite sides.

The other kinds of powerful radio sources are identified with star-like objects which are called quasi-stellar radio sources or shortly as QUASARS. The quasars are found to have radial velocities of recession equal to 90 per cent of the velocity of light. Consequently they are placed at large distance exceeding the distances of the most distant galaxies. The cause of the tremendous energy output of these star-like objects is still a mystery. However, their large distances make them important for solving the problems of cosmology discussed in the next section.

(*c*) *The expanding universe* : One

important feature of the galaxies is that they all appear to move away from us. The speed of recession 'v' increases proportionately with their distance 'r'. We can write $v = Hr$ which is known as Hubble's law. 'H' is known as Hubble's constant, it is found to have a value of 16000m/sec for every million light years. It is with the application of Hubble's law that we find the distances of objects like radio galaxies and quasars. Hubble's law indicates that the universe of galaxies is expanding. Based on these facts of observation several cosmological theories about the origin, and evolution of the universe are put forward.

According to the 'big bang' theory, all matter in the universe was concentrated in a very dense and hot primeval fireball at the beginning. An explosion occurred at about 20 billion years ago and since then the matter in the universe is moving away in the form of galaxies. Using the value of the Hubble's constant we find that the velocity of expansion equals the velocity of light at a distance of 20 billion light years. Since the light from the galaxies at that distance can never reach us it is called the boundary of the observable universe. On account of the continuous expansion more and more galaxies will go beyond this boundary and they will be lost. Consequently the number of galaxies per unit volume will go on decreasing and ultimately we will have an empty universe.

If the total mass of the universe is more than a certain value the expansion may be stopped by its gravitational pull and the universe may contract again. Thus we may have alternate expansion and contraction giving rise to a pulsating universe. There is also a third possibility suggested by T. Gold and F. Hoyle in U.K. They postulate that new galaxies are continuously being created out of empty space to fill up the gap caused by the galaxies which leave the observable part of the universe. This is known as the 'steady state theory.'

The presently available observations of galaxies and quasars are not sufficient to decide which of three theories of cosmology is the correct one. However, one piece of independent observation is against the steady state theory. It is observed that radio radiation of a few millimetre wavelength is coming from all parts of the sky. This background radiation has a temperature of 3K and it is believed to be the remnant of the original very hot radiation of 10^{10} K which existed at the time of the big bang. But the question whether the universe will continue to expand for ever or whether it is a pulsating universe is still not settled. We need to know the average density of the universe much better before a final answer can be obtained.

Table 1

PHYSICAL PROPERTIES OF THE OBJECTS IN THE SOLAR SYSTEM

Object	Period of revolution in years	'a' in A.U.	Radius R (Earth)	Rotation period	Mass M(Earth)	Mean density (kg/m×10⁻³)	g g(Earth)	Temperature (Degree C)	Albedo	Surface pressure in atmospheres	Atmospheric chemical composition.	No. of satellites etc.
Moon	—	—	0.27	d 27.32	0.0123	3.34	0.170	+110 (day) —150 (night)	0-07	0	Vacuum	—
Mercury	0.241	0.387	0.38	d 58.6	0.056	5.4	0.367	+340 (day) —120 (night)	0.06	0	Vacuum	0
Venus	0.615	0.723	0.96	d 243.0 (retrograde)	0.815	5.1	0.886	+480 (surface) +40 (clouds)	0.85	100	Co_2 (95%)	0
Earth	1.000	1.000	1.00	$23^h56^m.1$	1.000	5.52	1.000	+45 (Equator) —50 (Poles)	0.40	1	N_2(80%) O_2(20%)	1
Mars	1.881	1.524	0.53	$24^h27^m\pi4$	0.107	3.97	0.383	+30 (Eq.) —130 (Poles)	0.15	0.006	CO_2(97%) N_2(3%)	2
Ceres (Largest Asteroid)	4.603	2.767	0.055	90^h05^m	0.0001	3.34	0.18	—	0.07	0	Nil	++
Jupiter	11.864	5.203	11.23	9^h50^m5	317.9	1.33	2.522	—140(Clouds)	0.45	—	H_2, He, CH_4, NH_3	14
Saturn	29.46	9.540	9.41	10^h14^m	95.2	0.70	1.074	—175(Clouds)	0.61	,,	H_2, He, CH_4, NH_3	10+ring
Uranus	84.01	19.18	3.98	10^h49^m (retrograde)	14.6	1.33	0.922	—220(Clouds)	0.35	,,	H_2, He, CH_4	5+ring
Neptune	164.1	30.07	3.88	15h	17.2	1.66	1.435	—230(Clouds)	0.35	,,	H_2, He, CH_4	2
Pluto	247	39.44	0.5	d 6.39	0.11	4.9	0.440	—240 ?	0.14	?	?	Nil
Halley's comet	76.2	17.8	+	—	10^{-13} 10^{18}g		—	—	—	—	—	—

Notes : 1 year=365.257 days ; 1 A.U.=1.496 X 10^8 Km ; R (Earth)=6378 Km ; M (Earth)=5.977×10^{27} g ; g (Earth)=9.82 m/sec^2;
+Comet Nucleus diameter 10 Km, head 10000 Km ; ++ Total number of asteroids>1600.
d=day h=hour m=minute

Table 2

PROPERTIES OF STARS OF VARIOUS SPECIAL TYPES

Spectral type	Colour	Example	Description of Spectrum	Temperature	M / M (Sun)	R / R (Sun)	Absolute Magnitude Mv	L / L (Sun)
O	Very Blue	θOrionis	Lines of ionised Helium	35000	40	20	—6.0	10^5
B	Blue	Spica (Chitra)	Lines of neutral Helium	20000	15	7	—3.1	10^4
A	White	Sirius A (Vyadha)	Balmer lines of Hydrogen	9500	2.3	1.8	+1.4	25
F	Green	Procyon	Lines of Hydrogen and ionised metals	7000	1.4	1.2	+2.7	3
G	Yellow	Sun (Surya)	Lines of ionised and natural metals	5800	1.0	1.0	+4.8	1
K	Orange	Σ Eridani	Lines of neutral metals and molecular bands	4500	0.7	0.8	+6.1	0.4
M	Red	Kruger 60	Molecular bands of Tio	3500	0.3	0.4	+11.8	1/40
M	Red	Betelgeuse (Ardra)	Tio bands	3000	20	220	—6	5×10^4
A	White	Sirius B	Broad lines of hydrogen	9500	1.0	1/50	+11.5	1/200

Exercises

15.1 (a) Give a list of various constituents of the universe.

(b) How does the science of astronomy differ from other sciences like Physics ?

15.2 (a) Describe and compare the various kinds of telescopes used for astronomical observations.

(b) The lens of our eye has a diameter of 8 mm. How much fainter objects can be seen through a telescope of 120 cm aperture as compared to the faintest naked eye stars.

15.3 (a) The distances of the satellites of Mars are 25″ for Phobos and 62″ for Deimos at mean opposition when Mars is 0.524 A.U. for earth. Calculate distances of the two satellites for Mars in astronomical units and in metres.

(b) If the period of revolution is 0.319 day for Phobos and 1.262 day for Deimos use equation (15.3) to find the mass of Mars in units of the earth's mass which is 5.977×10^{24} kg.

15.4 (a) Explain why some bodies in the solar system have atmospheres and others do not.

(b) Discuss the possibility of life on the various planets in the solar system.

15.5 Period of revolution of the moon is 27.32 days and its mean distance from earth is 384,400 km. Use equation (15.2) to calculate the sum of the masses of the earth and moon. Further, using the known fact that the centre of mass of the earth-moon system lies at 4.75×10^6m from earth's centre, calculate the mass of each.

15.6 (a) Calculate the value of the solar constant at the distanse of Jupiter which is 5.2 A.U. away from the sun. Hence calculate the temperature of a black body on Jupiter by Stefan's law.

(b) Explain how the sun produces its energy.

15.7 (a) From the data given in columns 4 and 6 of Table 1, calculate and verify the entries about the mean density and acceleration due to gravity given in columns 7 and 8.

(b) A man weighs 68 kg on earth. Find his weight on the various objects in the solar system.

15.8 From the data in Table 2, calculate the mean densities and g-values on the surfaces of various types of stars.

15.9 (a) The two components of Procyon have a separation of 4".55 and period of 40.6 years. If the distance of the binary is 11.3 light years, calculate the sum of the masses of the two stars in solar units. If the ratio of their masses is 3:1 find the mass of eachs tar.

(b) If both components of Procyon have the same spectra while one component is 15000 times brighter than the other, calculate the ratio of their radii.

15.10 Narrate the complete life history of a star.

15.11 Describe the structure and contentsof the Milky Way.

15.12 (a) Give the properties of normal galaxies and explain how the radio galaxies and quasars differ from them.

(b) Briefly discuss the three main cosmological theories of the universe.

UNIT 16

Solids

16.1 Structure of Solids

Man has used solids from a very ancient period when he was dwelling in caves but it is only in this century that he has gained a satisfactory understanding of the way constituent atoms arrange themselves in a solid. You have come across commonly occuring crystals of sugar, rocksalt, mica, quartz and some gems. The crystals have symmetrical shapes and flat faces which meet at characteristic angles. Most of the solids, which do not appear to be crystalline to the naked eye, are actually found to be aggregates of small crystals when examined under a microscope. Questions like why diamond is hard, what makes copper a good conductor of electricity and iron a magnet, naturally occur to us. The first step towards answering such questions and understanding the nature and behaviour of solids is the knowledge of the structure of solids i.e. the knowledge about the arrangement of atoms in a solid.

The diffraction of X-rays by crystals demonstrated directly that a crystal is a periodic arrangement of atoms or molecules. We had already noted while discussing diffraction of waves that a periodic array of atoms in three dimensions would readily diffract X-rays whose wavelength is of the same order as the interatomic distance. The X-ray diffraction has been successfully used to probe the structure of variety of crystals ranging from simple materials like sodium chloride and copper to some very complex materials like proteins.

To describe the structure of a crystal we do not have to give positions of all the atoms in the crystal. The regularity and the orderliness which is there in the atomic arrangement makes this problem quite simple. The arrangement of atoms in a crystal can be described by choosing a small group of atoms and imagining them to be located in a cell. Repeating this unit cell along three definite directions in space provides the atomic positions in a crystal. We illustrate this in Fig. 16.1 by a two dimensional example for the sake of simplicity even though strictly speaking there does not exist, in the real world, a two dimensional crystal. Here the repeating unit is chosen to consist of two atoms represented by white and black circles. This is only a schematic representation and is not intended to show the shape and size of atoms faithfully. The choice of the unit cell is quite arbitrary. Two possible choices are shown in Fig. 16.1. The knowledge of the location of atoms in the unit cell contains complete information regarding the structure of the

crystal. This introduces an extreme simplification in the study of crystals because instead of considering the whole crystal, one has to consider only the unit cell.

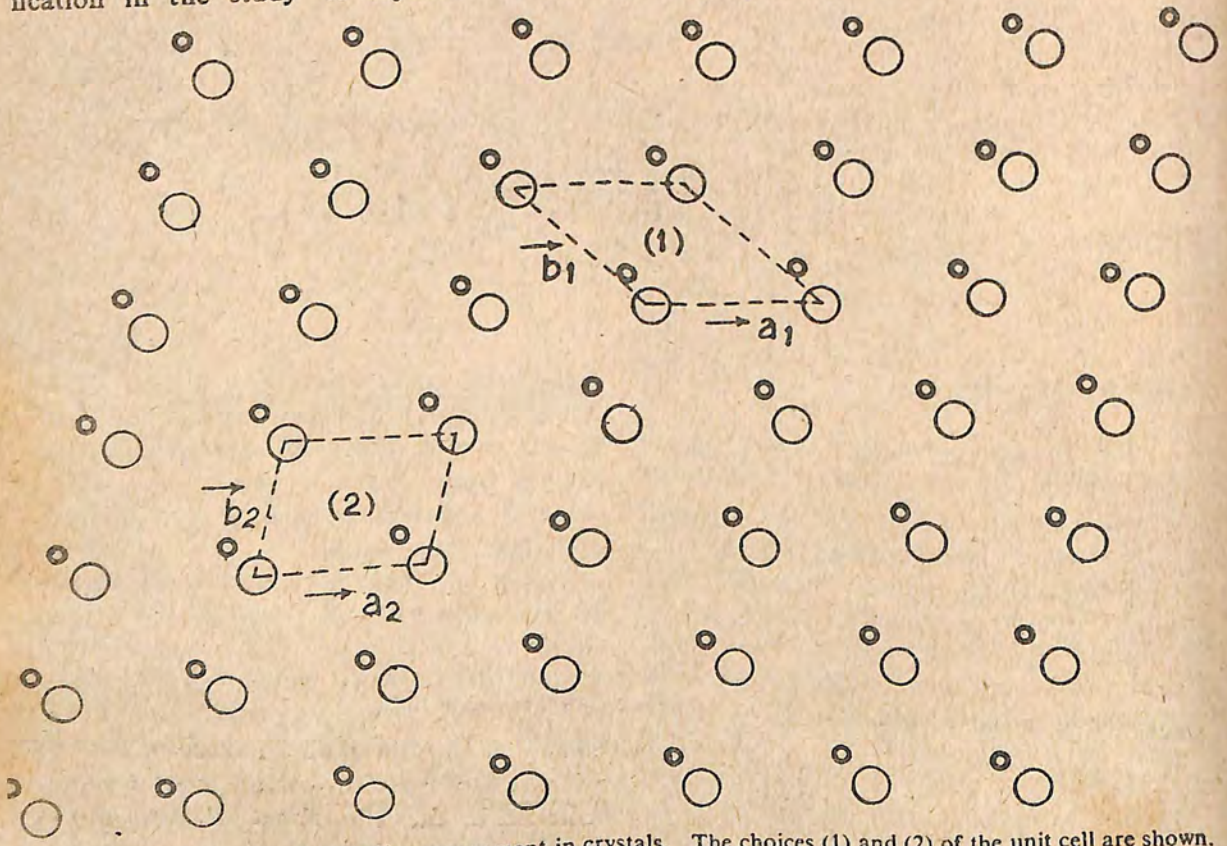


Fig. 16.1. The ordered periodic arrangement in crystals. The choices (1) and (2) of the unit cell are shown.

16.2 Crystal Lattice

The study of the geometry arrangement of atoms in a crystal becomes simple if we replace the crystal by a lattice. A lattice is a purely geometrical concept whereas crystal is a real thing. We explain the concept of the lattice with the help of the two dimensional structure as shown in Fig. 16.1. There we have a group of two atoms repeating themselves according to a pattern. If we associate a point with this group and replace all the groups of atoms by points placed at corresponding positions, we get a collection of points (Fig. 16.2). The array of these points is known as the lattice. A lattice is thus a regular arrangement of points in space. The crystal structure is obtained by attaching the group of atoms to lattice points. The lattice points are given by the tips of lattice vectors :

$$\vec{l} = n_1 \vec{a} + n_2 \vec{b} \qquad \ldots(16.1)$$

with some arbitrary origin O. $\vec{a}$ and $\vec{b}$ are *translation* vectors, and are shown in Fig. 16.2 for the two choices (which are indicated there by the subscripts to $\vec{a}$ and $\vec{b}$). n_1 and n_2 are arbitrary integers.

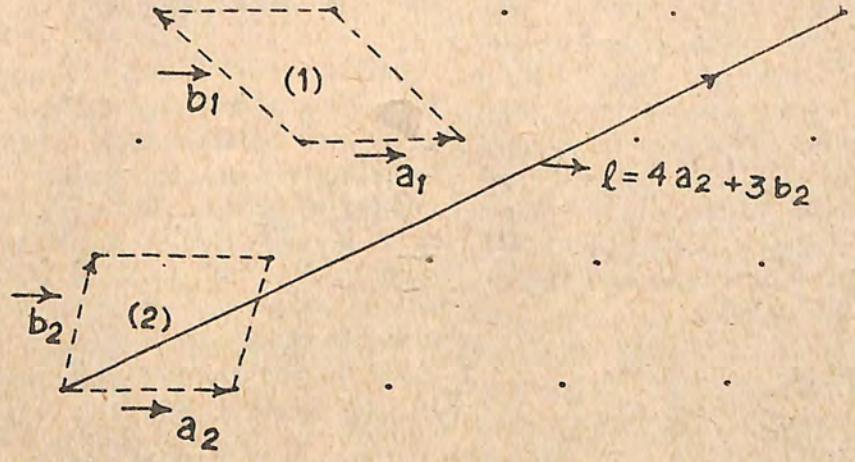


Fig. 16.2. The lattice corresponding to crystal structure shown in figure 16.1.

The Eq. 16.1 can be generalized to three dimensions where the lattice points of a three dimensional lattice would be given by

$$\vec{l} = n_1 \vec{a} + n_2 \vec{b} + n_3 \vec{c} \quad \ldots(16.2)$$

16.3 Amorphous Solids

We have seen above that the regular periodicity in the arrangement of atoms is the characteristic feature of *crystalline solids*. There is no such orderliness found in the arrangement of atoms in amorphous solids. Glass is an example of an amorphous solid. In fact, glass is used as a synonym for amorphous solid. Like liquids the amorphous solids have a disordered arrangement of constituents (molecules or atoms). The main difference in liquids and amorphous solids is that in liquids molecules are free to move within the volume of the liquid whereas in the amorphous solids the atoms are rigidly fixed.

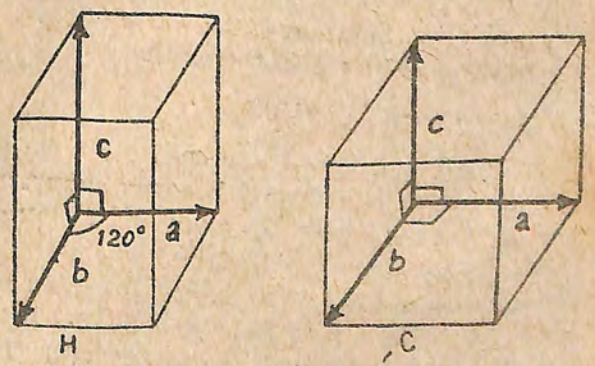


Fig 16.3. Crystal systems : H—hexagonal C—cubic.

The glass does not have a well defined freezing point. We know that when a liquid crystallizes it does so at a definite freezing point at which the heat of solidification (latent heat of freezing) is released. This corresponds to the lowering of the energy of the system due to the ordering of the atoms of the system to the regular arrangement in a crystal. In glasses the freezing into a crystalline state is prevented either due to high viscosity of the liquid or due to the fast rate of cooling. A glass, therefore, retains random disordered arrangement similar to that of a liquid.

16.4 Crystal Systems

We have seen in Fig. 16.2 that a variety of unit cells can be chosen for a crystal lattice. All crystal lattices can be grouped into seven crystal systems according to the seven different types of unit cells we can choose for them.

The unit cells for two of these seven systems "the cubic and the hexagonal" are shown in Fig. 16.3. Unit cell is specified in terms of the translation vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$

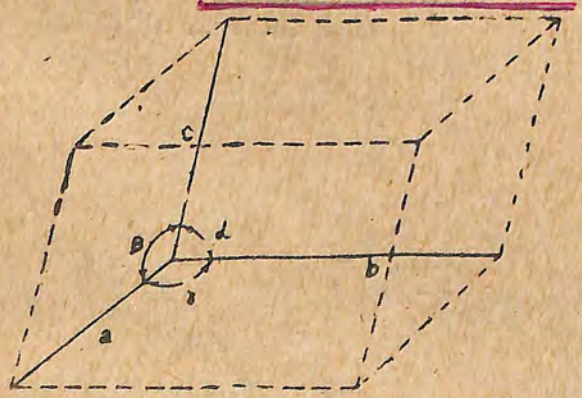


Fig. 16.4. A unit cell defined by vectors $\vec{a}$, $\vec{b}$, $\vec{c}$. The angles α, β and γ are shown, for example α is the angle between $\vec{b}$ and $\vec{c}$.

and the angles α, β, and γ, between them. The notation used for the angles is explained in Fig. 16.4.

16.5 Close Packed Structures

Many crystals found in nature, specially those of common metals and inert gases (the inert gases also crystalize at low enough temperatures) can be viewed in terms of packing of spheres, where the atoms in the solid are likened to hard impenetrable spheres. The two ways of close packing spheres (atoms) in three dimensions are discussed here. If we try to pack spheres as closely as possible in a plane, there is only one way shown in Fig. 16.5, where spheres

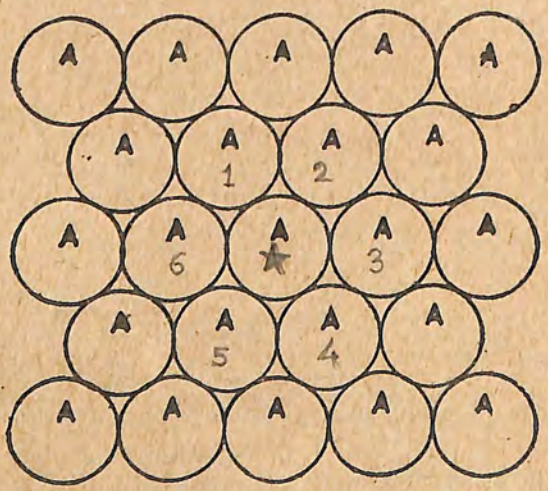


Fig. 16.5. The close packing of spheres in a layer.

are sitting at positions marked, A, forming an equilateral-triangular array. Here each sphere in the plane is surrounded by six spheres which just touch it. A second layer of spheres over this layer is placed so that each sphere (B) of this layer settles over a depression (centered at the point indicated by +) formed by three spheres in the first layer as shown in Fig. 16.6. A third layer can now be added in two *distinct* ways. In

one, the spheres in this layer are directly over the corresponding spheres in the first layer and we call this sequence of these three layers ABA. In the second arrangement

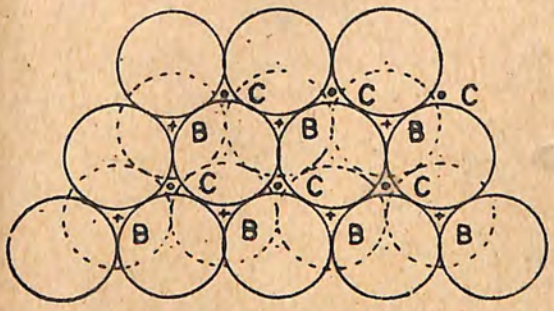


Fig. 16.6. The second layer of spheres is shown dashed.

the spheres in the third layer lie directly over the hollows marked C of the first layer. We call this sequence ABC. We note from Fig. 16.6 that hollows C are the ones which were left out when we placed spheres in the second layers at hollows marked B. The continuation of the above two ways of stacking spheres in space, ABABAB..... and ABCABCABC....., leads to two types of the close packed structures known as *hexagonal close packed* (hcp) structure and *face centred cubic* (fcc) structure respectively. A typical example of hcp structure is magnesium. Aluminium, copper, silver, gold, nickel, platinum and many other metals possess fcc structure. The position of atoms in a hcp structure is shown in Fig. 16.7. The unit cell for the structure has been outlined with heavy lines. A similar figure could be drawn for the fcc structure.

The above are the only two ways of arranging atoms in close packed structures. Compared to these, other structures are loosely packed. We see that in close packed structures there are twelve nearest neighbours to each atom, six in its plane, three in the plane above and three in the plane below. This number of nearest neighbours known as the *coordination* number is less for structures which are not close packed.

16.6 Forms of Binding in Solids

The properties of a solid clearly depend on the kind of binding that exists between atoms. The forces which bind atoms together in solids are all electrostatic in origin. The strength and the type of the binding is determined by the distribution of the electrons in the outermost shells in the atoms

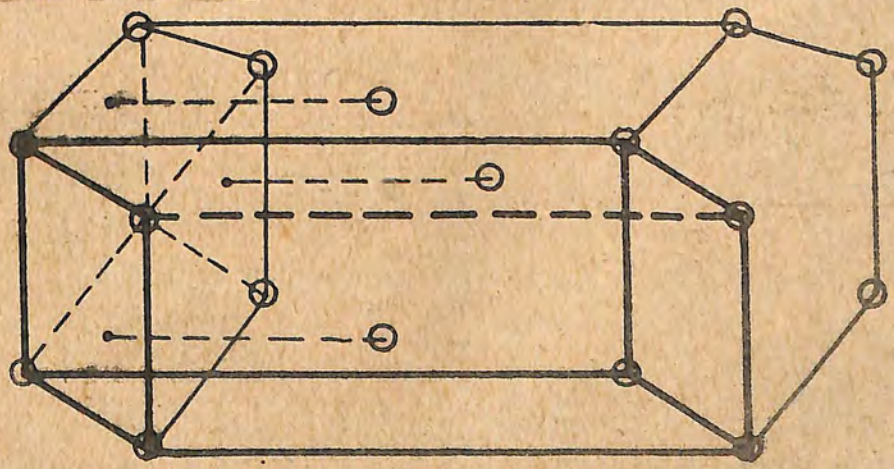


Fig. 16.7. Atomic positions in the hcp structure.

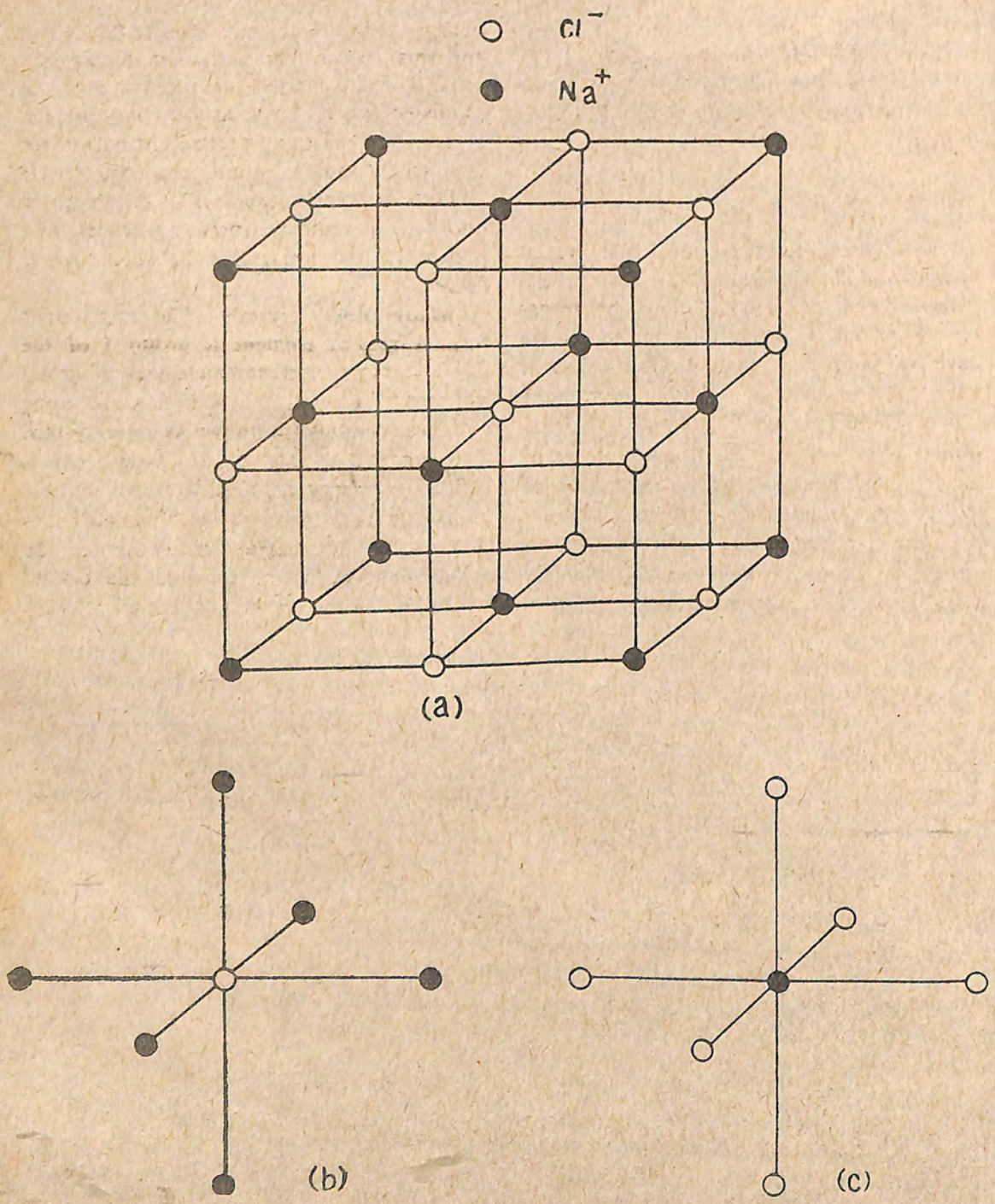


Fig. 16.8. Arrangement of ions in NaCl crystal. The figures (b) and (c) show nearest neighbours of a Cl^- ion and a Na^+ ion respectively.

involved. The binding in crystals is conveniently divided into four types : ionic, covalent, metallic and Van der Waals. We shall discuss each of these types of binding separately.

The Ionic Binding (1)

If we bring together equal number of sodium and chlorine atoms, it is easy for the valence (3s) electron of sodium to be transferred to the 3p shell of chlorine and we have Na^+ and Cl^- ions. Both these ions have stable closed shell structures ($Na^+ = 1s^2\ 2s^2\ 2p^6$ and $Cl^- = 1s^2\ 2s^2\ 2p^6 3s^2\ 3p^6$) similar to those of inert gas atoms In common salt (sodium chloride) the positive ions of sodium (Na^+) and the negative ions of chlorine (Cl^-) are arranged in the manner shown in Fig. 16.8. In this arrangement the nearest neighbours of any ion are ions of opposite charge (Fig. 16.8b and Fig. 16.8c). It is for this reason that attractive forces between opposite charges predominate over the repulsive forces between the like charges. If the effective force remained attractive for any separation of ions, the ions would collapse into each other. This is prevented due to the repulsive force that comes into play when the separation between ions is decreased.

Many other crystals, like compounds formed from an element of group I of the periodic table and some element of group VII, are held together by this ionic binding and these crystals are known as ionic crystals. The cohesive energy of an ionic crystal is defined as energy required to transform the ionic crystal into a system of isolated ions (Fig. 16.9). The higher the value of the cohesive energy the more stable the crystal would be. Ionic crystals have high cohesive

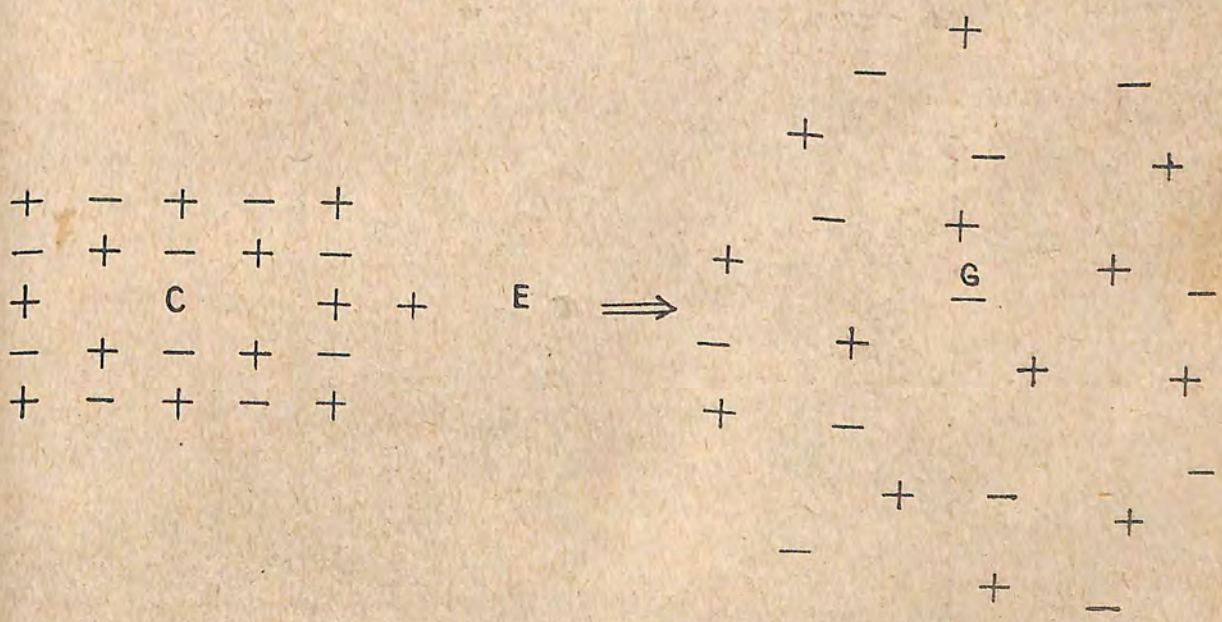


Fig. 16.9. Schemetic illustration of cohesive energy of an ionic crystal. C—crystal, E—cohesive energy, G—gas of isolated ions.

energy, therefore the melting points of ionic crystals are also high. Because all the electrons in these crystals are bound to that ion core, there are no free electrons to carry electric current, such solids are therefore poor conductors of electricity.

Covalent Binding (2)

In this class of crystals the constituent atoms are bonded together by sharing of electrons with neighbours. Group IV crystals like diamond, silicon, germanium have covalent bonds. For example, in diamond, a carbon atom is involved in four covalent bonds with its four nearest neighbours which lie at the vertices of a tetrahadron centred at the atom under consideration (Fig. 16.10). This is a loose packed structure with coordination number four. The four outer electrons of the carbon atom can join with electrons of adjacent four carbons give a strong directional bond. Each covalent bond is formed by two electrons and each carbon atom sharing the bond contributes one electron to it. The electron charge density for these two electrons participating in the bond is concentrated along the bond in between the two atoms. This is shown schematically in Fig. 16.11. The neighbouring bonds repel each other and the four bonds at tetrahedral angle (109.5°) are directed towards the four vertices of the tetrahedron. The directional nature of the covalent bonds does not allow the crystal to assume the close packed structure. Crystals with covalent bonds are very hard. The materials are difficult to deform due to directional nature of the bond.

Many solids are partly ionic and partly covalent. The ionic character progressively increases in the sequence of solids Ge (IV group element)→ GaAs (III—V compound)

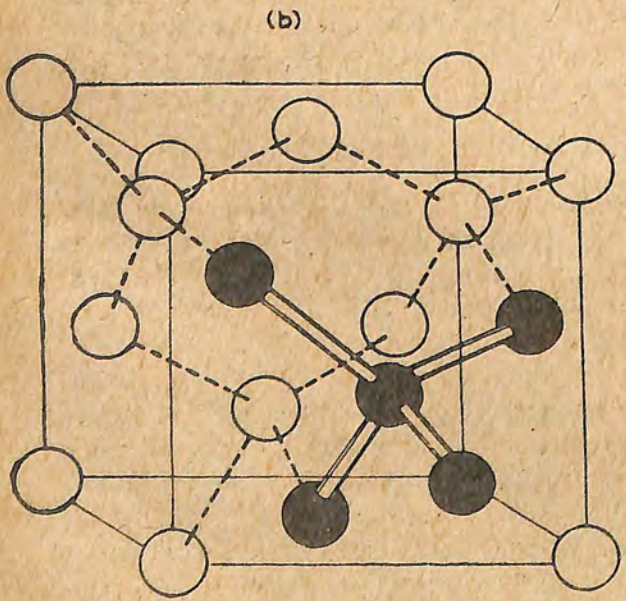


Fig. 16.10 The unit cell of the diamond showing tetrahedral bonding in the crystal. One of the tetrahedral units have been marked out by denoting the carbon atom involved in this unit by filled circles and the bonds by double lines.

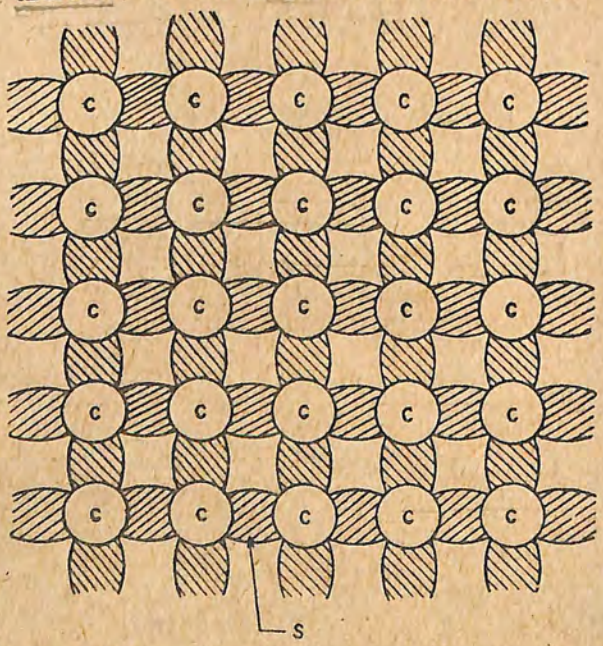


Fig. 16.11. Highly schemetic representation of the electron distribution in covalent bonding in diamond. S—shared electron.

→ZnSe (II—VI compound) →CuBr (I—VII compound). For example, in ZnSe all the electrons are not shared, Zn and Se are partly charged due to permanent transfer of some electrons from Zn to Se so that the crystal is partly ionic.

Metallic Binding (3)

When a large number of atoms having only a few weakly bound electrons in the outermost subshell, are brought together, a metallic crystal is formed where the electrons are detached from the atoms and these electrons move freely through the lattice of positive ions. A metal is therefore a regular array of ions embedded in a nearly uniform distribution of free electrons pervading the whole crystal (Fig. 16.12). This freedom of movement of electrons in the solid makes them behave in some ways like a gas of electrons. The behaviour of the electron gas is quite different from the ordinary gas of atoms because here the constituents are negatively charged electrons compressed into a much smaller space than the ordinary gas, say in copper there are more than 10^{22} electrons in a cubic centimetre. The electrostatic attraction between the electrons and ions dominates over the mutual repulsion between electrons and between ions, and this is what holds the metal together. The electron gas, therefore, serves the purpose of glue in metallic binding. There is no directionality in the binding of the metal. It is for this reason that quite a few metals have close packed structure. It is also because of this non-directionality that the metals can be alloyed with one another.

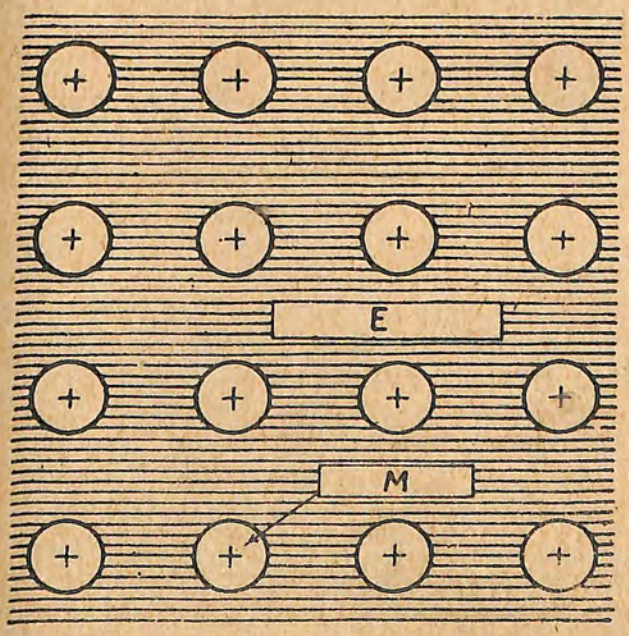


Fig. 16.12. A metal can be regarded as an array of ions pervaded by electron gas. E-electron gas, M-metal ion.

The freedom of the movement of electrons through the crystal is responsible for the high electrical conductivity of metals. The binding energy of alkali metals, such as sodium, is smaller than in ionic crystals such as NaCl, and such metals are quite deformable.

Van der Waals Binding (4)

Inert gas atoms have completely filled outer shells and these should have virtually no interaction with other identical inert gas atoms. It is found that at low enough temperatures (Ne : 2.4K, Ar : 84K, Kr : 117K) inert gases also condense into solids, indicating that there exists a weak attractive interaction between the atoms. If there were no such interactions to bind the atoms the system would not condense. In Unit 12, it has been mentioned that inert gas atoms and molecules interact through Van der Waals forces. The bonding in inert gas crystals is due to Van der Waals bonds. In addition to inert gas crystals, the molecular crystals like paraffin, benzene etc., are also held together by the Van der Waals interac-

tion. The Van der Waals bond is considerably weaker than the ionic, covalent and metallic bonds. Crystals with Van der Waals binding would therefore have low melting points.

From our discussion of the four types of binding in solids we have seen that the interatomic bonds in covalent crystals are directional, and generally the metallic, ionic and Van der Waals bonds are nondirectional. The solids in which the atoms are bonded by nondirectional bonds tend to crystallize in close packed structures. The covalently bonded solids crystallize in structures which conform to the requisite angles between the directed bonds.

16.7 Electrons in Solids

Energy Bands in Solids

In order to understand physical properties of solids we need understand the electronic structure of solids as discussed below.

From our knowledge of the electronic structure of hydrogen atoms, we can say that the energies of the electrons in an atom cannot assume arbitrary values but only some definite prescribed values. Let us illustrate the problem by taking sodium atom which has 11 electrons (Fig. 16.13) arranged in the four shells ($1s^2\ 2s^2\ 2p^6\ 3s^1$). The three inner shells (1s, 2s and 2p) are completely filled and there is only one electron in the outermost (valence) shell 3s (which can accomodate 2 electrons when completely filled). The energy levels which are occupied by electrons in the various (1s, 2s, 2p, 3s) shells as well as the next unoccupied shell are shown in Fig. 16.14(a). Fig. 16.14(b) illustrates how energy bands arise in solid sodium. As a large number of atoms are brought together, they begin to influence each other. Because of the interatomic interaction the electrons in outer shells of the atoms are forced to have energies different from those in atoms. Since a crystal would consist of a large number of the atoms (typically about 10^{23}), we have an enormously large number of levels to deal with. Each atomic level will split into about 10^{23} levels. Because of the such large number of levels in a small energy range (say a few eV at the most) one regards these energy levels to be continuously distributed in energy within a range of energy. We call this an allowed band of energy. These *energy bands* are separated by regions of energy in which no levels exist. This forbidden region of energy is known as band gap or energy gap. In Fig. 16.14(b) we show how the bands are formed when atoms are imagined to come closer and closer from infinite separation in a crystal of sodium. According to the Pauli exclusion principle one level can accomodate only two electrons. If we start filling up nearly continuously distributed energy levels in the bands, we find that the

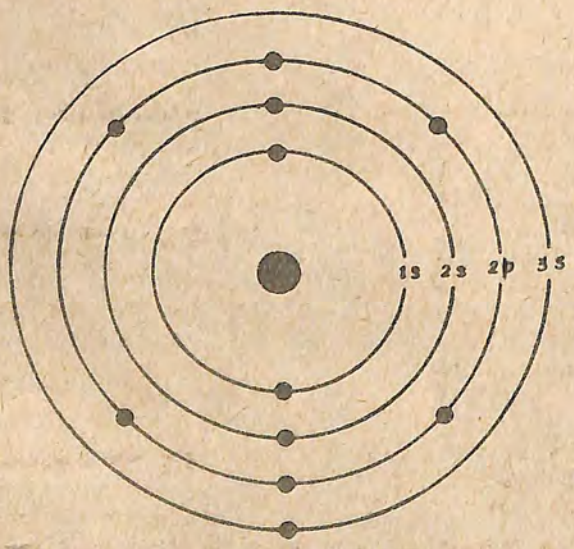


Fig. 16.13. A schematic representation of the electron configuration of the sodium atom.

bond formed out of 3s level will be half filled. This is shown in the Fig. 16.14(b) by filling up the occupied portion of the band. The equilibrium separation of atoms in the solid sodium is 3.6 A°. The conventional energy level diagram for solid sodium is obtained from energy levels as they appear at interatomic separation of 3.6 A° in Fig. 16.14(b). Fig. 16.15 shows the energy band of the sodium crystal corresponding to the lattice spacing of 3.6 A° (shown by the dashed vertical line on Fig. 16.14(b)

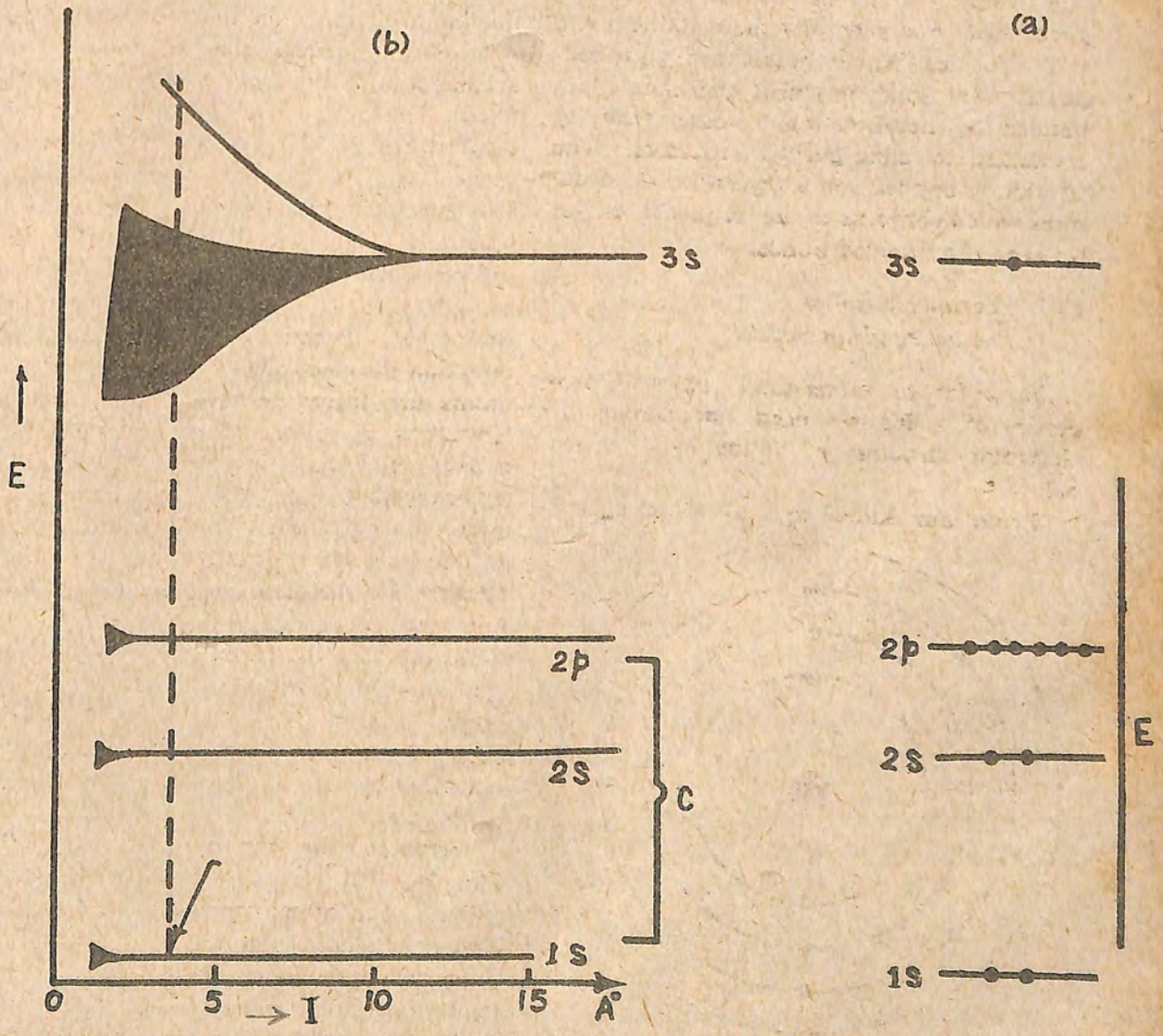


Fig. 16.14. (a) Some energy levels of Na atom (b) This schematically shows the formation of energy bands from energy levels of the sodium atoms which are made to approach each other. The dashed line indicates the interatomic seperation in solid sodium. E-energy, I-interatomic spacing, C-Ion core states not split in the crystal.

Metals, Insulators and Semiconductors

The band picture for energy states of electrons developed above can be used to classify solids into metals, insulators and semiconductors.

Let us consider three possibilities regarding the spacing and the occupancy of the energy bands as shown in Fig. 16.16. In case of a metal the topmost band with electrons is not completely filled by them. This partially occupied band is known as the conduction band. In the case of an insulator the uppermost energy band having electrons is completely filled and the next one completely empty but separated from the filled band by a large energy gap. In the case of a semiconductor this gap is

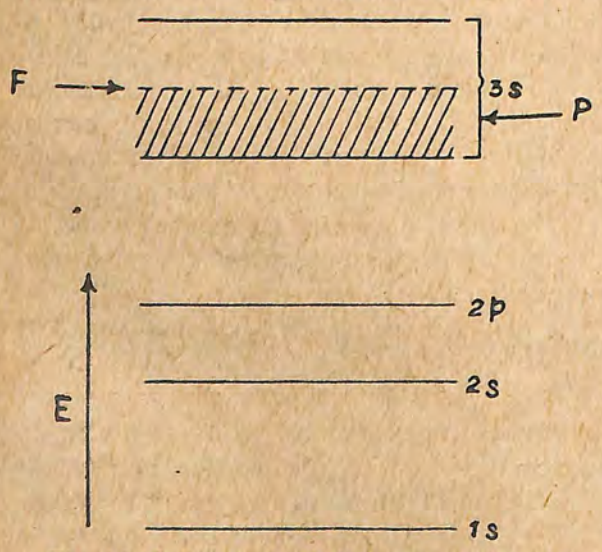


Fig. 16.15. Energy bands for solid sodium. Core levels 1s, 2s and 2p are not broadened by interatomic interaction.
F = fermi energy, P-partially filled (conduction) band, E-energy.

comparatively small (say less than about 2eV) silicon with a band gap of about 1eV is a semiconductor where-as the diamond with a band gap of 7eV is an insulator. The completely filled band is known as the valence band. The next band which is empty but which has electrons excited from filled states (in valence band or due to impurities) is known as the conduction band.

The electric current flowing through a solid is caused by the momentum gained by electrons due to force acting on them in an applied electric field. This process is not possible if the bands in a solid are completely full because the exclusion principle does not permit any redistribution of electrons in a filled band. Electrons, therefore, cannot gain any energy from the field and the current cannot flow. The solid is an insulator. In a metal, the conduction band is partially filled. The highest energy level in the conduction band occupied by electrons at absolute zero is called the Fermi level. The energy corresponding to the Fermi level is the *Fermi energy*. In the presence of an applied field the electrons in a metal can be excited to the empty energy levels immediately above the Fermi level and some of the electrons now are accelerated in the direction of field.

At finite temperatures, it is possible for some electrons in the valence band of an insulator with a relatively small gap, to gain enough thermal energy to jump over the energy gap into the conduction band. The probability of the promotion of electrons from valence to conduction band increases exponentially with temperature and decreases also exponentially with the magnitude of the band gap. When a few electrons are excited thermally into the conduction band they leave behind an equal number of empty states near the top of the valence band. After promotion of electrons the valence band is not completely full and can contribute to conduction. It is more convenient

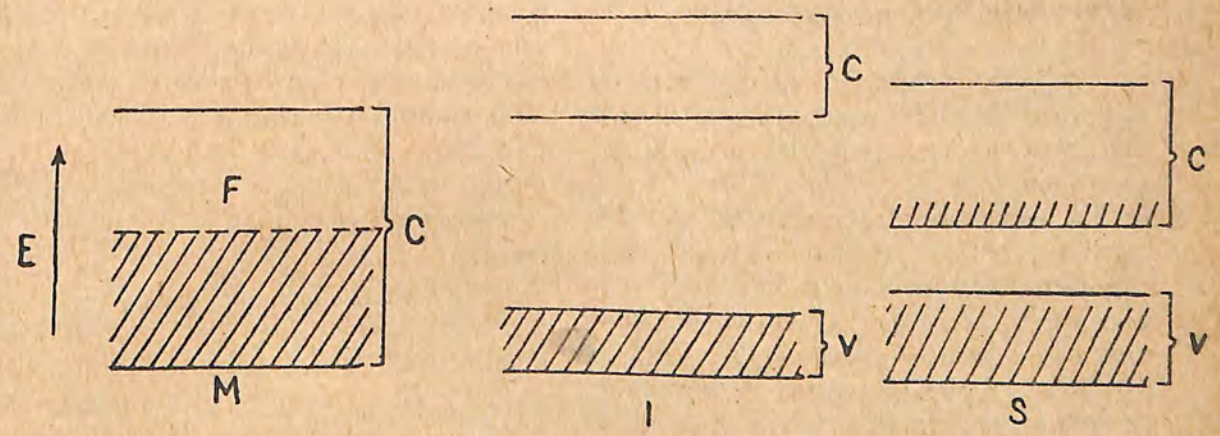


Fig. 16.16. Filling of the energy bands in a metal, an insulator and a semiconductor, shaded areas indicate energy regions occupied by electrons. E-energy, M-metal, I-insulator, S-semiconductor, F-fermi level, G-energy gap, C-conduction band, V-Valence band.

to talk in terms of the few empty states in the valence band rather than the more numerous occupied states. These empty states or 'holes' act as if they were positively charged carriers. In semiconductors we shall have electrons in the conduction band and an equal number of holes in the valence band as charge carriers for the current flow.

16.8 Semiconductors

Intrinsic Semiconductors

We have seen above that semiconductors are materials with a small band gap between the valence and the conduction band. The study of semiconductors has led to dramatic developments in electrical devices such as transistors and has brought about a technological revolution in electronics. (See Unit 20).

If we are dealing with a pure semiconductor, its electronic properties are dominated by electrons excited from the valence band (Fig. 16 16). The electrical conductivity in these pure semiconductors is called intrinsic conductivity. In a pure semiconductor at absolute zero the valence band is completely full and there are no electrons at all in the conduction band. At ordinary temperatures appreciable number of electrons would be excited across the energy gap because of the energy of thermal motion of the atoms of the semiconductors. If the energy gap is narrow enough or temperature is high enough, a large number of electrons would be promoted to the conduction band.

We can discuss this intrinsic conductivity in terms of the *bond* model illustrated in Fig. 16.11. Let us take the example of the semiconductor silicon. We had seen in article 16.6, while discussing the covalent bond that the four valence electrons belonging to each silicon atom are used in forming the covalent bonds with its four nearest neighbours. These electrons forming the bonds are not affected when an electric field is applied to the crystal. To say the same thing in terms of the band picture one would say that the electrons forming the bonds completely fill the valence band of the crystal. Substantial amount of energy is needed to release these electrons from

their bonding state to a free state in the conduction band and this energy could be imparted thermally due to the heat motion of atoms at any finite temperature. Therefore at any temperature, other than absolute zero, there would always be a finite number of free electrons to carry current and the number of such carriers will rise very rapidly with the increase of the temperature. In terms of the language of the band theory this corresponds to thermal excitation of electrons from the valence band to the conduction band. A schematic representation of the intrinsic semiconductors is given in Fig. 16.17. When an electron is broken apart from a bond, a vacancy is created in the bond. This absence of an electron from the bond is called the hole. An electron from a neighbouring bond can move into this hole and exchange places with it. The motion of the hole in a particular direction is, therefore, equivalent to the motion of an electron in the opposite direction. In this manner in the presence of a field both the electrons and holes can move and give rise to an electric current. The electrons would move towards the positive potential and the holes in the opposite direction.

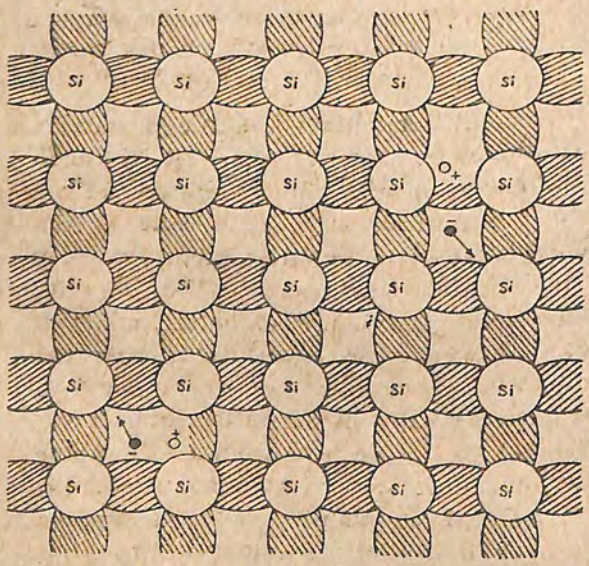


Fig. 16.17. In a pure crystal of the silicon, the electron breaks away due to thermal energy from the covalent bond providing us with an electron free to conduct electricity. This electron leaves behind a vacant state called hole which also move from one bond to another. In the language of the bond theory the electron is promoted thermally from the valence band to conduction band leaving behind a hole.

Extrinsic Semiconductors

Addition of impurity atoms, whose valence is different from four, also introduce electrons and holes into the host crystal and drastically influence the electrical properties of the semiconductor. Atoms of an element belong to group V of the periodic table, say phosphorous or arsenic, can be substitutionally dissolved in Si or Ge. Only four of the five electrons of the impurity atoms participate in forming the bonds and the fifth electron from each impurity is almost free to conduct electricity. We say almost free because a very weak attraction of this electron towards the positively charged impurity atoms is always there. The coulomb attraction between the positively charged impurity ion and the electron is weakened by the (dielectric constant of the) medium. This would mean that these electrons from impurity atoms would have energies a little less than the energies of the mobile (free) electrons belonging to the conduction band. In the band picture the energy states for these electrons would lie a little below the bottom of the conduction band as shown in Fig. 16.18. This electron can be readily freed by energy of heat motion. The impurity atoms which can give up an electron in this manner are called donors. A deliberate addition of

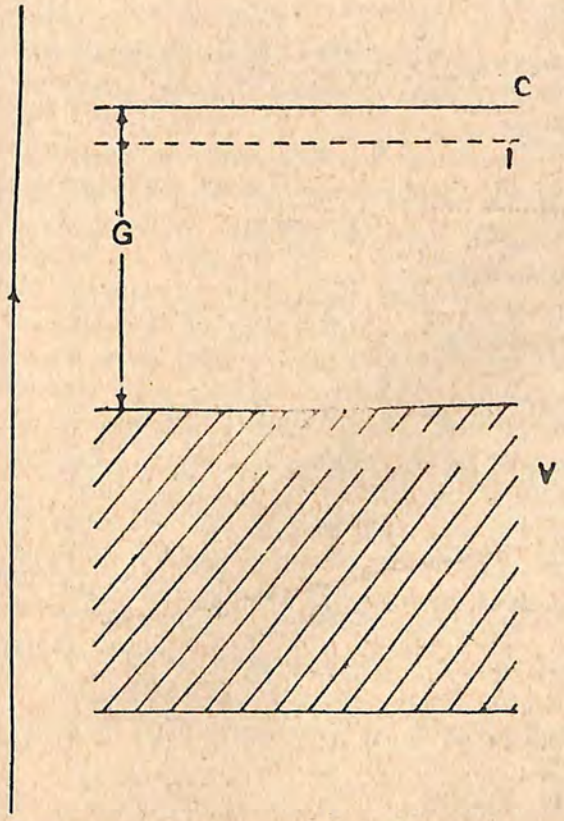


Fig. 16.18. Impurity states due to donors in an extrinsic semiconductor.
C-conduction band, I-impurity levels for donors, V-valence band, G-energy gap.

impurities is called doping. A semiconductor like Si or Ge doped with pentavalent impurities like phosphorous or arsenic is called an n-type semiconductor because it has an excess of negatively charged electrons. n-type silicon is shown in Fig 16.19. The electrical conductivity of the n-type semiconductor will be determined by number of free electrons in the conduction band and the ease with which they move in the crystal is called the mobility of electrons.

If an impurity atom belonging to group

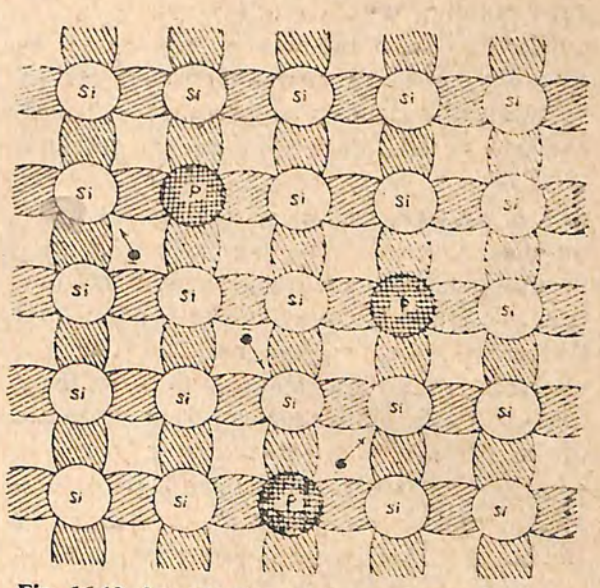


Fig. 16.19. Introduction of electron as carrier in a n-type semiconductor obtained by addition of phosphorus (P) to silicon (Si)

III; boron, aluminium, galium and indium, substitutes an atom of the host semiconductor crystal like Ge or Si, then there would be a vacant place or a hole in the bond because the impurity atom has only three valence electrons. The impurity atoms that give rise to such holes are known as acceptors. The hole formed near the impurity may be filled up by an electron from one of the neighbouring bonds and would leave a hole in this bond. This hole may again be filled up by another neighbouring electron. In this manner the hole moves through the crystal. The movement of this deficiency of a negative charge can be regarded as the movement of a positive electric charge. One, therefore, associates a positive charge with the hole. A semiconductor crystal doped with an acceptor is called a p-type semiconductor because it has an excess of positively charged holes. The formatian of holes in p-type silicon doped with boron is

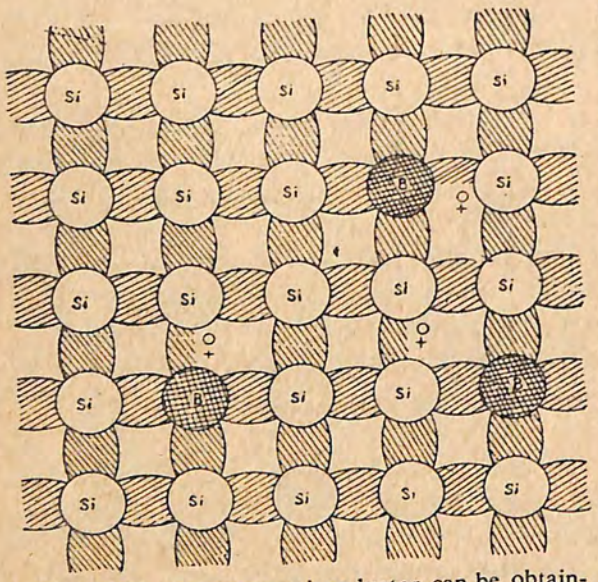


Fig. 16.20. A p-type semiconductor can be obtained by doping Silicon(Si) with traces of boron(B). Boron with only three valence electrons to contribute, creates a hole into which a nearby electron can move.

shown in Fig. 16.20. Impurity concentrations as small as 100 parts per million lead to an extrinsic behaviour for the semiconductor.

The above discussion tells us that the electrical conductivity of both the intrinsic and extrinsic semiconductors would increase with temperature. In the case of metals the conductivity decreases with increase of temperature. Irrespective of the manner we introduce electrons or holes in a semiconductor, whether by heating or by adding impurities, these carriers are nearly as mobile as the electrons in a metal. But the number of carriers in a semiconductor is much less compared to the electrons in metal, therefore the conductivity of semiconductors is smaller than that of metals.

Exercises

16.1 Calculate the $\frac{c}{a}$ ratio for the hexagonal close packing of spheres with the help of Fig. 16.7.

16.2 Show that the fraction of total volume filled by spheres for hcp structure is 0.74.

16.3 Draw the unit cell for the simple cubic (Sc) structure. Find the number of nearest neighbours. (6)

16.4 Copper has fcc structure with lattice parameter a=3.6 A°. Estimate the atomic radius of Cu. (1.27 A°)

16.5 Separate out from the following list the materials which form a close packed structure Cu, Zn, Si, Ge, Al, Fe, Ni, Pb, Cd, Ne, Ar.

16.6 How many lattice points are there in the cubic unit cell of the bcc structure ? (2)

16.7 What type of bonding do you expect in the following solids :
MgO, NaI, SiC, GiTe, C_6H_6, BN, CuNi.

16.8 Why are metallic solids always opaque, covalent solids sometimes opaque and ionic crystals hardly ever opaque to visible radiation ?

16.9 Why does metallic binding occur mostly with atoms having small number of valence electrons ?

16.10 Of the ionic, covalent, metallic and Van der Waals solids which class is most likely to produce an insulator, a conductur, a semiconductor ?

16.11 What elements other than As and Sb can be used as an impurity with germanium to form an n-type semiconductor ? What elements other than gallium and indium can be used to form a p-type semiconductor ?

16.12 Indicate the nature of binding in the solids with the following description :

(a) Reflects light in the visible, electrical resistivity increases with temperature, melting point 1000°C.

(b) Reflects light in the visible, electrical resistivity decreases with temperature and melting point above 1000°C.

(c) Transmits light in the visible, conducts electricity only at high temperatures ; and

(d) Transmits light in the visible but very low melting point.

16.13 One alloys copper with zinc, will the band width of the conduction band increase or decrease ?

16.14 What factors determine the electrical conductivity of a metal ?

UNIT 17

Properties of Materials

17.1 Elastic Properties of Solids

A lot of our technological development has been possible through our understanding of the behaviour of materials. Our search for materials possessing desired mechanical properties has met with partial success. Methods have been developed by which we can improve the mechanical properties like hardness, strength, brittleness, etc., of alloys ceramics and polymers; but we still have not reached the stage where we can make materials having a desired value for a mechanical property.

The study of mechanical properties of a solid in general deals with the deformations due to forces applied on the materials. A solid tries to maintain its shape under the influence of small external forces and recovers its original shape and size when the forces are removed. This property of the solid is known as elasticity.

Young's Modulus

If we stretch a material with certain force in a particular direction, then the material gets elongated along the direction of the force as shown in Fig. 17.1.

As a result of external forces there would arise internal forces within the material and stress T refers to these internal forces. The

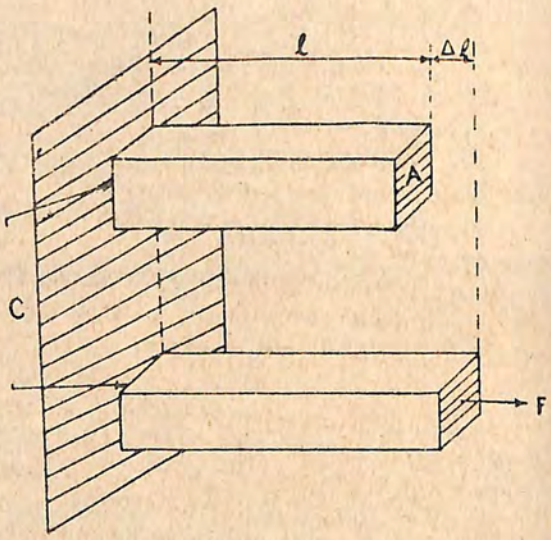


Fig. 17.1 Elongation of a bar under a tensile stress. C—clamped end.

Stress T is defined as force per unit cross-sectional area. The *strain*, e, is a measure of the deformation of the material.

When a body is strecthed the strain is defined as the elongation per unit length. If a force F acts on a bar of length ℓ and cross sectional area A, and produces an elongation $\Delta\ell$, then

$$\text{Stress} = T = \frac{F}{A}$$

$$\text{and strain} = e = \frac{\Delta\ell}{\ell}$$

Results of experiments show that the relationship between the stress and strain depends on the material. The stress-strain relationship for a typical material like copper is schematically shown in the Fig. 17.2. In the limit of very small elongations the stress is proportional to strain.

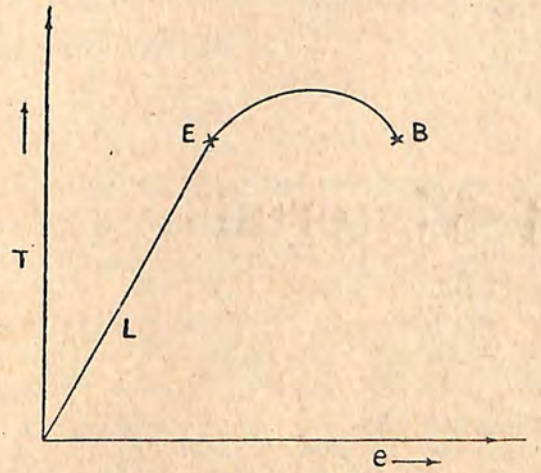


Fig. 17.2 Stress [T] and strain [e] graphs for a material like copper which undergoes deformation before breaking. L—linear region, E—elastic limit, B—point at which the material breaks.

$$T = Ye \qquad \text{.....(17.1)}$$

The constant of proportionality Y is known as *Young's modulus*. Y depends only on the nature of the material. The linear relationship 17.1 between stress and strain is known *as Hooke's law*. A low value of the Young's modulus means a material which stretches a lot for small stresses.

EXAMPLE 17.1

How much will a 3.0 m long copper wire elongate if a weight of 10 kg is suspended from one end and the other end is fixed. The diameter of the wire is 0.4 mm.

Solution

We shall take Y for copper to be 1×10^{11} N/m^2 and g = 9.8 m/sec^2.

$$\therefore \text{ stress} = \frac{mg}{\pi r^2} = \frac{10 \times 9.8 \text{ N}}{\pi (2\times 10^{-4})^2 \text{m}^2} = 7.8 \times 10^8 \text{ N/m}^2$$

$$\text{strain} = \frac{\Delta \ell}{\ell} = \frac{\Delta \ell}{3}$$

$$\text{Hence} \quad \frac{\Delta \ell}{3} \times (1\times 10^{11}) = 7.8\times 10^8$$

$$\text{or } \Delta \ell = 2.34 \text{ cm.}$$

Non-elastic Behaviour

If a specimen returns to its original shape and size on removal of the stress, the deformation is said to be within the elastic limit. Beyond the elastic limit, the applied forces produce plastic deformation and the body will not recover its orginal shape even if the stress is withdrawn i.e. the material undergoes a permanent deformation. If we increase the strain beyond the elastic limit the material will break at some value of strain. A material which breaks in this way is called ductile. A ductile material is therefore permanently distorted before breaking. Most of the metals are ductile.

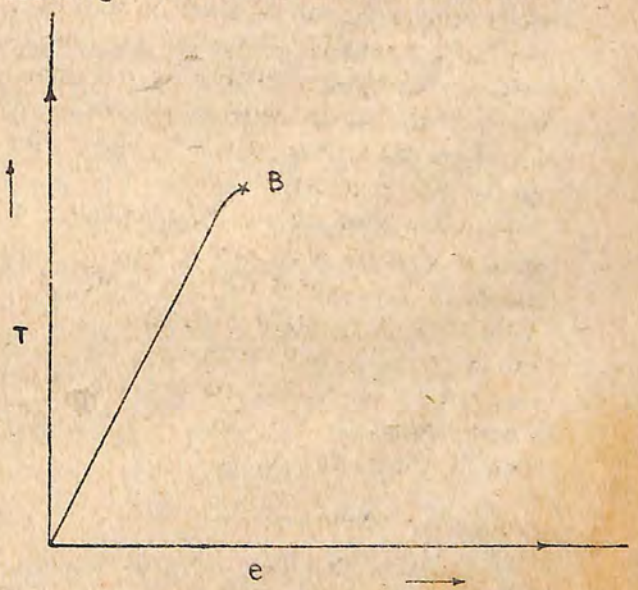


Fig. 17.3 Stress [T]—Strain [e] curve for a brittle material like cast iron. B—breaking point.

Some metals like cast iron or materials like glasses and ceramics break quite suddenly as soon as the stress-strain curve starts deviating from a straight line (Fig. 17.3). Such materials are called *brittle*. Brittle materials are difficult to form into different shapes where as ductile materials can be shaped.

In the above we have only considered the tensile (or compressive) force which stretches (or compresses) a solid. We can also deform a solid by hydrostatic stresses that is by pressures or tensions acting uniformly over the whole surface of the body or else by twisting forces which we will discuss below.

Bulk Modulus

If a body is subjected to uniform pressure acting normal to its entire surface it is said to be subjected to 'hydrostatic' pressure' (Fig. 17.4). A body submerged in liquid will feel such a pressure. Even though such pressures may be exerted by any other method, the pressure is called hydrostatic. As shown in the Figure if the pressure (which is the force per unit area), P, on the body is increased its shape remains unchanged, but the volume changes. The strain is defined as the ratio of the change in volume $\triangle V$ to the original volume. The stress within the body is equal to the change in pressure $\triangle P$,

$\therefore$ Stress = K $\times$ strain

where the constant of proportionality K is known as the bulk modulus of elasticity. Therefore

$$K = - \frac{\triangle P}{\triangle V/V} = - V \frac{\triangle P}{\triangle V} \quad \ldots 17.2$$

The minus sign is included because an increase in the hydrostatic pressure leads to decrease in the volume, that is if $\triangle P$ is positive, $\triangle V$ is negative.

The reciprocal of the bulk modulus is known as the compressibility.

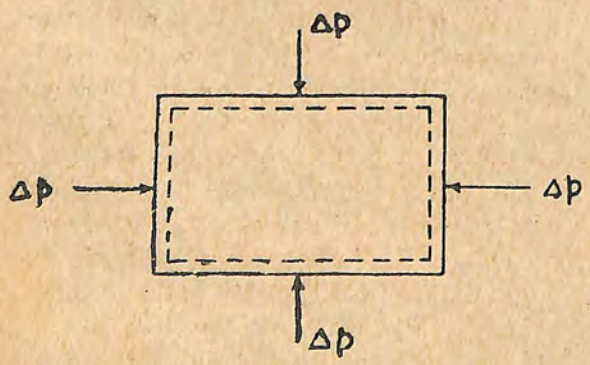


Fig. 17.4. Hydrostatic pressures $\triangle$ P causes a decrease in the volume.

Shear Modulus

If a pair of forces G act in opposite directions upon the upper and lower faces of a rectangular block, these forces shear it into a parallelopiped, turning the vertical end faces through small angle (Fig. 17.5). The shear stress is defined as the tangential force on a face divided by the area A of the face. The shear strain is the angle of twist. μ known as the shear modulus or coefficient of rigidity relates the stress and strain.

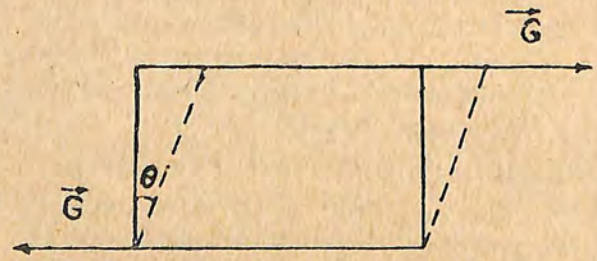


Fig. 17.5 The shear stress produces a change in shape without changing the volume.

stress $= \mu \times$ strain,

so that $\mu = \frac{G/A}{\theta}$ ... 17.3

17.2 Elastic Properties and Interatomic Forces

We had seen earlier that the atoms in a solid, interact with each other through interatomic forces. If we could plot the potential energy of interatomic interactions against interatomic seperation we would find that in a crystal the atoms are located at positions of minimum potential energy. The net force (which is the derivative of the potential) on an atom is zero at these potential minima. A tensile stress will increase the separation of atoms along the direction of the stress and the atoms are displaced from the potential minima to positions of higher energy due to the applied force. If the force is released the atoms will move back to their equilibrium positions. The displacement suffered by individual atoms for a particular tensile force will depend on the strength of the interatomic forces. Stronger the interatomic forces, smaller will be the displacements (strains) and higher would be the value of the young's modulus.

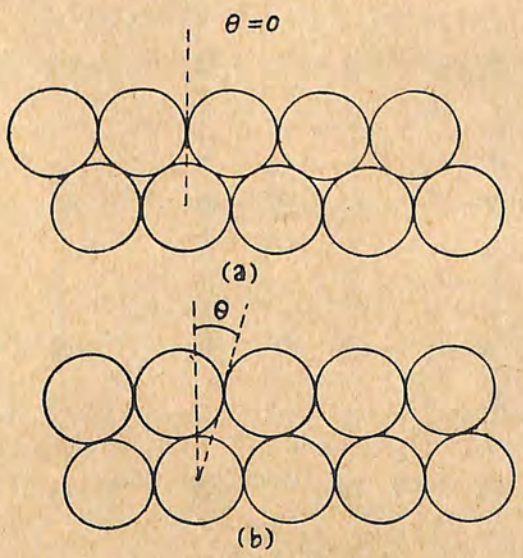


Fig. 17.6 (a) Two adjacent planes of atoms are shown in a stress free crystal.
(b) A shear strain makes these planes slide past each other.

In the case of shear stresses the distance between atoms remains the same, but neighbouring planes of atoms in a solid slide past each other. Figure 17.6 (a) show the initial location of atoms in two neighbouring planes, the location of atoms after the shear deformation is shown in Fig. 17.6 (b). This corresponds to the shear strain shown in the figure.

17.3 Thermal Properties of Solids

In our daily life we need materials which could store heat, materials which expand very little when heated, materials which are bad (or good) conductors of heat. We, therefore, need to understand thermal properties of materials. We discuss here qualitatively the heat capacity, the thermal expansion and the thermal conductivity of solids.

Heat Energy

Heat energy is identified with the motion of atoms in a solid, while describing the structure of solids we assumed that the atoms sit at the lattice sites. But an atom really does not remain fixed at a lattice site, but oscillates about it. A configuration of atoms at an instant is shown in Fig. 17.7. The thermal energy is stored in the vibratory motions of these atoms. Each atom experiences a potential energy of the form shown in Fig. 17.8. In the absence of any kinetic energy, the atoms will sit at their potential minima. This would be the situation at O K. At a higher temperature the atoms would execute oscillatory motion. In the neighbourhood of the minimum the poten-

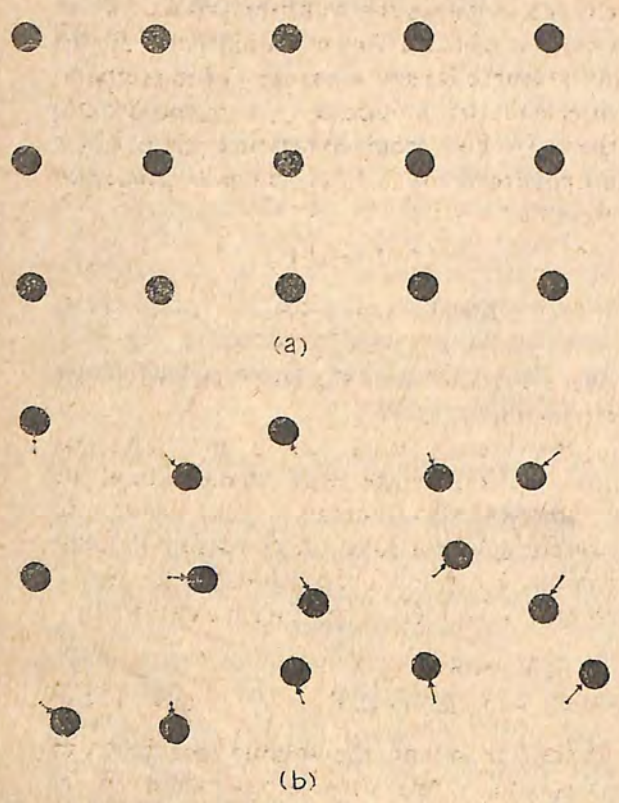


Fig. 17.7 (a) Idealised picture of atoms sitting at lattice sites ; (b) Instantaneous configuration of vibrating atoms in a solid.

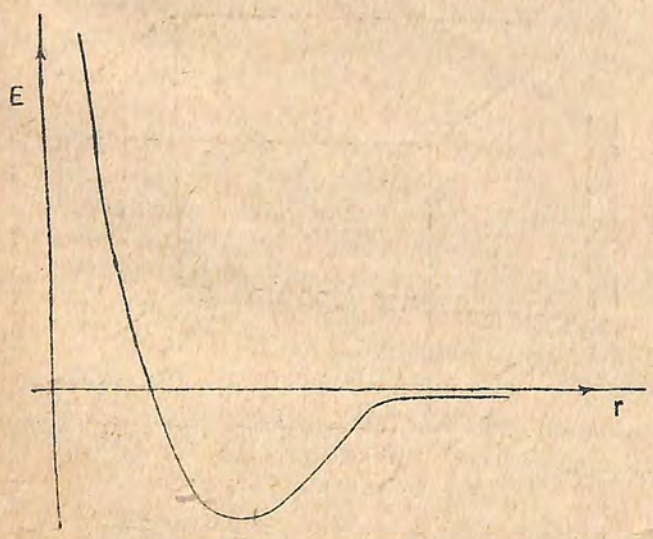


Fig. 17.8 Schematic form of interaction between atoms in a solid ; r—separation between atoms, E—potential energy.

tial function would be nearly parabolic. A parabolic potential energy means that the motion of the atoms would be simple harmonic. But an atom does not oscillate independently of other atoms. A displacement of an atom in a particular direction would influence the position of its neighbouring atoms and thus the motions of atoms are correlated. These facts, therefore, suggest a mechanical model for the vibration of atoms. The thermal energy of the solid is the energy of the vibrational motion of these atoms. In this model the solid may be regarded as a system of mass points linked by springs (Fig. 17.9). The motion of such a system would be quite complicated. This energy of the vibratory motion will be greater if the mass points have larger displacements from their equilibrium positions. Under ordinary conditions the amplitude of these atomic oscillations are of the order of 10^{-9} cm and the frequencies of the oscillations are about 10^{13} HZ.

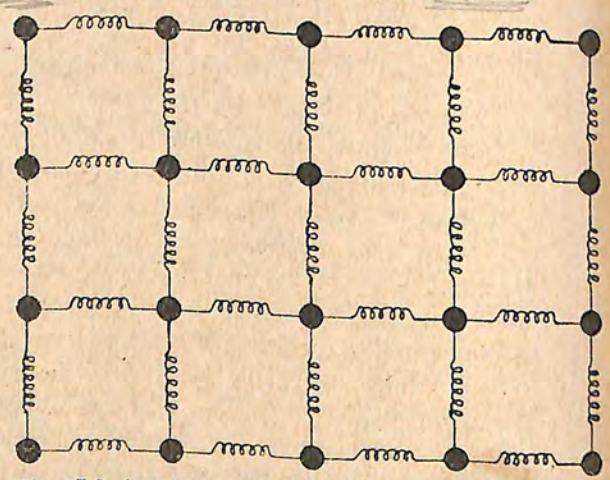


Fig. 17.9 A mechanical model for a vibrating crystal lattice.

Specific Heat

We have seen above that when heat energy is added to a solid, it is used up in exciting the vibratory motion of the atoms and the temperature of the solid rises. If the quantity of heat energy $\triangle Q$ supplied to a standard mass of a substance under certain

conditions leads to an increase in temperature ΔT, then we define $\Delta Q/\Delta T$ as the specifice heat of that mass. If one chooses the standard mass to be a mole (also called gram molecule) the specific heat is called the molar specific heat. The units of the molar specific heat are J mol^{-1} deg^{-1}. Experimental measurements show that near room temprature the value of specific heat

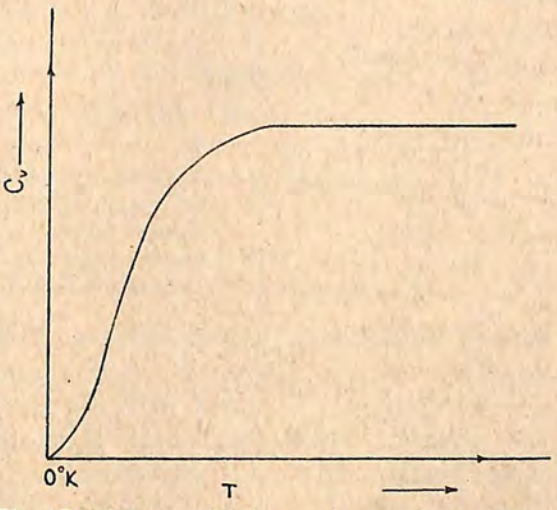


Fig. 17.10 The specific heat of a typical solid at constant volume C_v as a function of temperature, T.

of nearly all solids is close to 3R or 25 J mol^{-1} deg^{-1} where R is the gas constant. This is known as the Dulong and Petit rule. At lower temperatures the specific heat decreases rapidly as shown in Fig. 17.10.

It is not possible to explain the drop in the value of the specific heat with the lowering of temperature using a classical theory. The explanation by Einstein of the variation of the specific heat with the temperature as well as with the nature of the solid was one of the earliest successes of the quantum theory. We shall not discuss the details of it here.

17.4 Thermal Expansion

Nearly all solids expand on heating. The change in length (or width or thickness) of a solid is called a *linear expansion.* If the initial length is ℓ and a change ΔT in temperature leads to a change $\Delta \ell$ in the length, then we find from experience that $\Delta \ell$ is proportional to ΔT as well as ℓ. Hence we are able to write

$$\Delta \ell = \alpha \ell \Delta T$$

$$\text{so that} \quad \alpha = \frac{\Delta \ell}{\ell \Delta T} \qquad \ldots (17.4)$$

where α is known as the coefficient of linear expansion.

Because the length, width and thickness of a solid increase with temperature, its volume will also increase. The change in volume ΔV of a solid on increasing its temperature by ΔT is proportional to the initia volume V and T. Therefore we can write

$$\Delta V = \beta V \Delta T$$

$$\text{and} \quad \beta = \Delta V/V\Delta T \qquad \ldots (17.5)$$

where β represents the volume coefficient of expansion. We assume the solid to be isotropic for which the coefficient of linear expansion is the same along any direction. For example, if we take a rectangular block of the solid as shown in Fig. 17.11, the coefficient of linear expansion α will be the

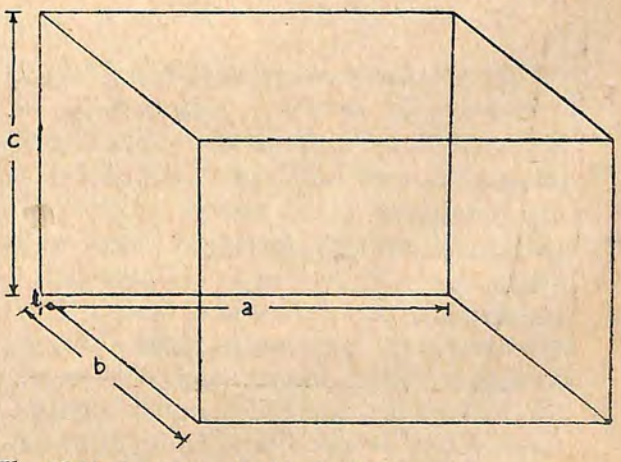


Fig. 17.11 A rectangular block of a material.

same for the length, the width and the thickness or for that matter, for any other length in this block. One can show that for an isotropic solid the volume coefficient is three times the linear coefficient of expansion. In order to show this, consider the rectangular block of Fig. 17.11 whose length, width and thickness are given by a. b and c respectively. The volume is therefore

$$V = abc$$

If we increase the temperature of the block by 1K, the changed length, width and thickness would become $a(1+\alpha)$, $b(1+\alpha)$ and $c(1+\alpha)$. Therefore the volume is

$$V + \triangle V = a(1+\alpha)\ b(1+\alpha)\ c\ (1+\alpha)$$
$$abc\ (1+3\alpha+3\alpha^2+\alpha^3)$$

Because the coefficient of thermal expansion α is always a small quantity, we can neglect $3\alpha^2 + \alpha^3$ and write

$$V + \triangle V = abc + abc\ 3\alpha$$

Because $V = abc$, we can write for change in volume for one degree rise in temperature to be

$$\triangle V = 3\alpha V$$

This by definition is equal to βV, therefore

$$\beta = 3\alpha \qquad \ldots (17.6)$$

In terms of our atomic picture, thermal expansion is due to an increase in the average distance between atoms as the temperature increases. As we increase the temperature the heat energy available for atomic oscillations increases. This would mean that with the rise in temperature the amplitude of the ocillation of atoms would increase. If the interatomic potential energy is exactly symmetrical with respect to the point where the potential energy is minimum, we see from Fig. 17.12 that the

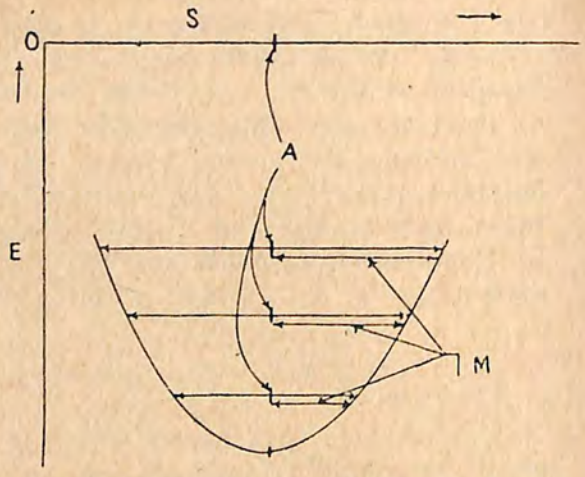


Fig. 17.12 Vibrational motion of atoms in a solid for a symmetrical interatomic potential energy (E). S—interatomic separation ; A—average inter-atomic separation ; M—amplitude of vibrational motion.

average interatomic distance between atoms does not change when the amplitude of atomic vibrations increases with the rise of temperature. The atom vibrates about the same average interatomic separation with a

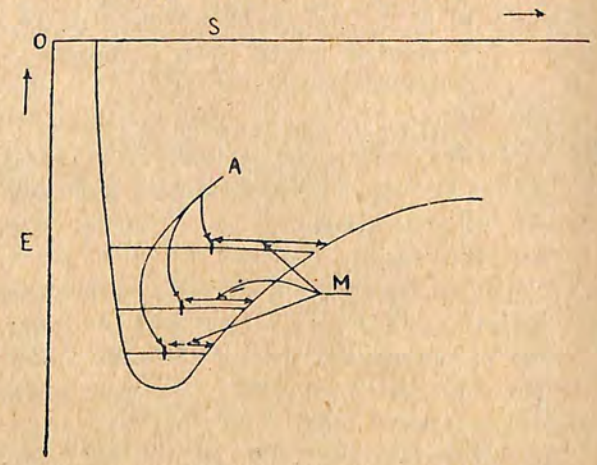


Fig. 17.13 The interatomic potential (E) for a real crystal is asymmetric. S—interatomic separation ; A—average interatomic separation ; M—amplitude of vibrational motion.

larger amplitude at a higher temperature. We see from Fig. 17.12 that thermal expansion would not exist if we use a symmetric potential.

The actual interatomic potential is some what asymmetric (not symmetrical). We have shown such an asymmetric potential in Fig. 17.13. We see in the figure that as the amplitude of oscillation increases with temperature, the average interatomic separation also increases. We can also see that the value of the thermal expansion coefficient will depend on the details of the shape of the potential energy function. In general the more asymmetric the potential, the larger would be the value of the thermal expansion coefficient.

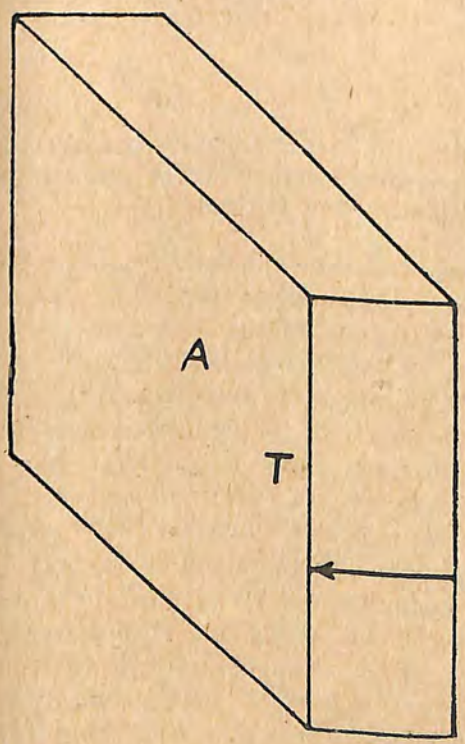


Fig. 17.14 Heat flow through a slab.

17.5 Thermal conductivity

If we take a piece of a solid and heat one end of it, then heat energy will flow from the hot end to the cold end. It is because of this reason that when we place one end of a metal rod over a flame, we feel the other end also heated up. The phenomenon of transfer of heat energy from one part of a solid to another part at a lower temperature is known as the conduction of heat.

Consider a slab of material of thickness Δx and cross sectional area A (Fig. 17.14). The two faces of the slab are at temperatures T and $T + \Delta T$. An amount of heat ΔQ will flow from the hotter face of the slab to the colder one in time Δt. Experimentally one finds that the rate of flow of heat through the slab, $\Delta Q/\Delta t$ is directly proportional to the temperature difference ΔT, and to area A, but is inversely proportional to the thickness Δx. Therefore

$$\frac{\Delta Q}{\Delta t} \propto A \frac{\Delta T}{\Delta x}$$

The constant of proportionality is represented by K called the thermal conductivity of the material. Thus the heat conduction equation becomes

$$\frac{\Delta Q}{\Delta t} = KA \frac{\Delta T}{\Delta x}$$

A material for which thermal conductivity K is large, is a good conductor like copper and aluminium. For thermal insulators like glass, wood and ice K has a low value. Usually metals are good conductors of heat and insulators are poor conductors

On the atomic scale a picture of thermal conductivity of an insulator can be presented as follows. When a part of a solid is heated, the atoms there start oscillating violently about their equilibrium position.

As mentioned, all atoms are bound to each other through interatomic interactions, therefore oscillation of an atom sets the neighbouring atoms into oscillation. The atoms at the hotter end will be vibrating with larger amplitudes than atoms at the colder end. An atom vibrating with larger amplitude will try to increase the amplitude of vibration of its neighbour which initially was vibrating with a little smaller amplitude and thus gives energy to it. This neighbour in turn gives energy in the same way to its neighbour and so on. Thus the atoms at the colder end will receive energy from atoms at the hotter end of the material.

In the case of metals heat energy can also be transported by electrons in addition to the mechanism discussed above for insulators. Since the electrons are quite mobile they are much more effective in transporting energy from hotter to colder parts of a solid compared to the mechanism given above. The thermal conductivity of metals is therefore, essentially determined by electrons.

In a metal the electrical conduction is due to the mobile electrons. Since the thermal conductivity of metals is also due to conduction electrons coefficient of the thermal conductivity of a metal is related to its electrical conductivity. In an insulator (which we shall also call dielectric) there are no conduction (or free) electrons and we may tend to think that the eiectric field will have no effect on the insulator. This is not the case. When a dielectric is placed in an electric field, if the molecules constituting the dieleectric have a permanent dipole moment, then the dipole moment of the molecules will tend to align themselves along the direction of the field as shown in Fig. 17.15. Thermal motion of molecules does not permit a complete alignment. In the absence of the field these dipoles were randomly oriented. Even if the constituent molecules do not possess a permanent dipole moment, the applied field will induce dipole moments in them. The induced moments would already be lined up along the direction of the field, because the induced dipole moment is along the direction of the field. Therefore, an electric field will polarize a dielectric.

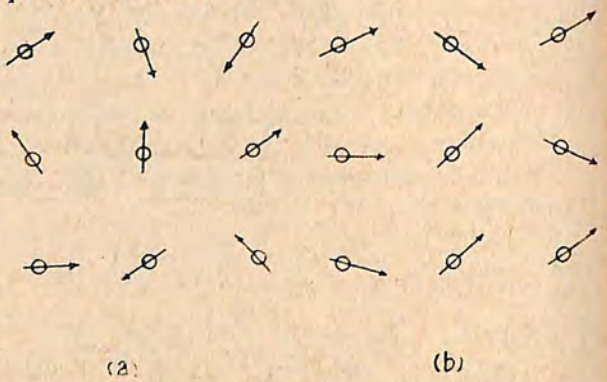
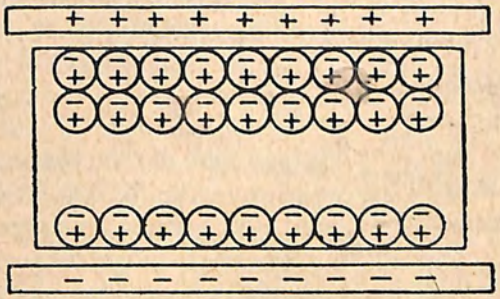


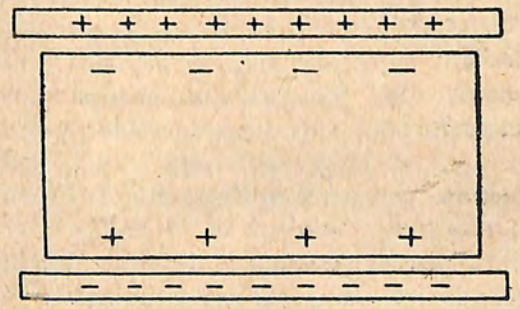
Fig. 17.15 Effect of an electric field on a dielectric whose constituent molecules possess permanent dipole moment.

If we place a slab of a dielectric between the plates of a parallel plate capacitor, the electric field will polarize the dielectric i.e. the negative charge of the molecule will be shifted slightly towards the positively charged plate and the positive charge towards the negatively charged plate (Fig. 17.16 a). These displacements are very small and are of the order of molcular diameters (10^{-10} m. In the bulk of the slab there will be no net charge but at the top surface in Fig. 17.16. (b) there would be an excess of negative charge and at the bottom surface there would be an excess positive charge due to polarization of the material. The charges in the capacitor plates set up an electric field E_o, the induced surface charges on the dielectric slab will set up a field—E′ in a

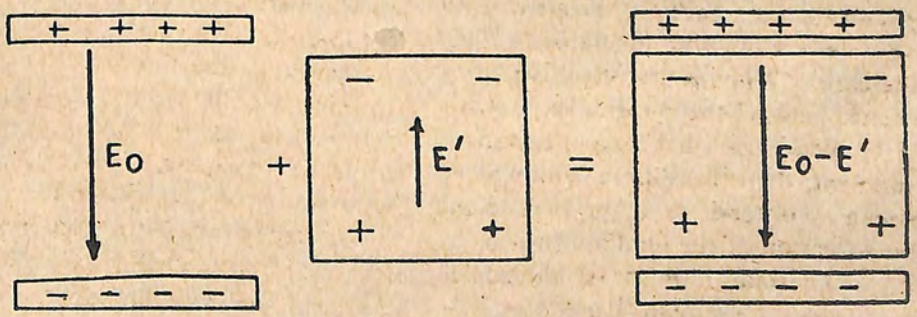
direction opposite to that of E_o. Therefore the resultant field, is E_o-E'. Thus when a dielectric is placed in an electric field, the surface charges are induced in the dielectric, which tend to reduce the original field. This is shown in Fig. 17.16 (c). The ratio $\frac{E_o}{E_o - E'}$ is known as the dielectric constant of the dielectric. If the dielectric consists of polar molecules, their orientation



(a)



(b)



(C)

Fig. 17.16 Capacitor with a slab of dielectric placed between the plates.

in the applied field would give a larger valve of E′, therefore a smaller E_0-E′. The dielectric constant of such a dielectric would be large.

17.6 Magnetic Properties of Materials

From the point of view of the magnetic behaviour we can classify the materials into three categories : 1. diamagnetic, 2. paramagnetic, and 3. ferromagnetic. In all materials except the ferromagnetic materials the magnetic effects are quite small. Ferromagnetism is shown by materials like iron, nickel, cobalt. Any magnetic field would be strongly influenced by the presence of ferromagnetic materials. A piece of diamagnetic material (e.g. bismuth) when placed in a non-uniform field will experience a force directed from stronger to the weaker part of the field. A piece of bismuth brought near the pole of a strong magnet will be repelled. This will be a weak effect. A paramagnetic matarial, salts of transition metals or rare earths like chrome alum placed in a non-uniform field will experience a force directed from weaker to the stronger part of the field. Therefore, a piece of aluminium placed near a strong magnet will feel a weak attraction. A ferromagnet will experience a relatively larger force directed from weaker to the stronger part of the field. The effect for a ferromagnet is so large compared to a paramagnet or a diamagnet that paramagnets and diamagnets are popularly called 'nonmagnets'.

The magnetic properties of matter are traced to motion of charges. The motion of an electron in an atom or a molecule would lead to magnetic effects. The spin of the electron is also a source of magnetism of the particle. It is not possible to present an accurate picture of magnetism without using quantum mechanics, because spin of an electron as well as its motion in a solid can be described accurately only by using quantum mechanics. We shall, therefore present models which will give qualitative picture of diamagnetism, paramagnetism and ferromagnetism.

Diamagnetism

We had seen earlier that in an atom the electrons move in circular orbits around the nucleus. The rotating electron will behave like a current loop. This current loop is equivalent to a magnetic moment equal to the product of the current and the area of loop (here the electron orbit). The direction of the moment is normal to the plane of the orbit (Fig. 17.17 a). In a material there will be a very large number of such electrons and the orbits of the electrons will be oriented randomly. The magnetic moments due to orbital motion of electrons cancel each other and there is no net magnetism, Fig. (a) and (b). If a magnetic field B is impressed from outside, the electron

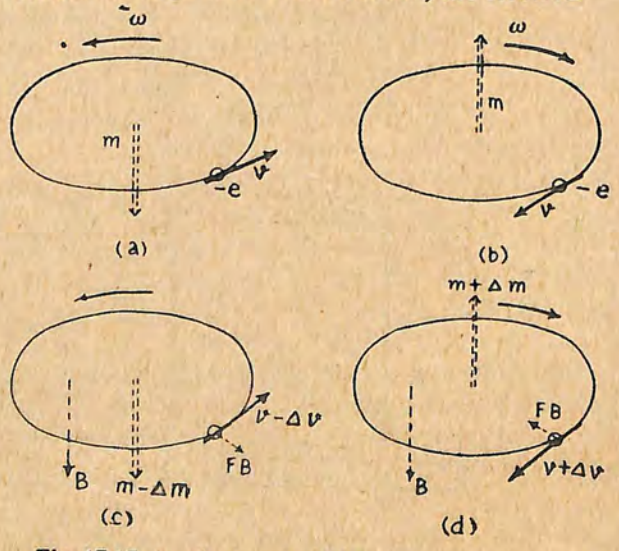


Fig. 17.17 An electron orbiting in an atom produces a moment.

experiences a radial force $\vec{F}_B = -e(\vec{v} \times \vec{B})$ which is at right angles to the direction of motion. Since the radius of the orbit remains constant, the angular velocity of the electron will decrease or increase depending on whether F_B acts radially outwards, Fig. (c) or inwards, Fig. (d). Thus the effect of applying the magnetic field is to decrease or increase the angular velocity of the orbiting electron. A reduced angular velocity in (c) would mean, a reduced current and therefore a reduced magnetic moment. Similarly the magnetic moment of the electron orbiting an opposite direction in (d) would increase the magnetic moment. The two electrons orbiting in the manner shown in Fig. (a) and (b) pair up to give a net zero moment in the absence of the magnetic field. In the presence of the field such a cancellation is not possible because the net effect of the two orbits (c) and (d) in Fig. 17.17 is to give a moment in a direction opposite to that of the impressed field. The moments due to all the electrons in a material in the presence of an external magnetic field will add up to a finite magnetization of the material in a direction opposite to that of the field. This accounts for diamagnetism of the material. This also shows that diamagnetism should be present in all materials because the orbital motion of electrons in atoms is a universal phenomena.

Paramagnetism

From the above discussion you might expect all materials to be diamagnetic. But experiments show that this is not the case. Many materials are paramagnetic because in them paramagnetism completely dominates the diamagnetism. The paramagnetism is due to the fact that in many solids, the ions or atoms in it have permanent magnetic moment. These moments come from the orbital motion of the electrons and from the magnetic moment intrinsic to each electron arising from its spin. We find that in atoms whose inner electron shells are not completely filled there is a net moment in them. This is the case for atom of transition elements like iron, nickel, cobalt, palladium and for rare earth atoms like gadolinium, etc.

If we place a magnetic moment $\vec{m}$ in a magnetic field $\vec{B}$, the field will try to align the moment along the direction of the field, because a moment pointing along the direction of $\vec{B}$ will have the lowest energy. Let us consider a solid containing N atoms each of which has a magnetic moment $\vec{m}$. The interactions between these moments is very weak and we can regard these moments to be independent of each other. In the absence of a magnetic field these atomic moment would be completely randomly oriented because of the thermal agitation of the atoms at any finite temperature. When we apply an external magnetic field, it will try to line up the magnetic moments along the direction of the magnetic field. This process of alignment along the direction of the magnetic field is hampered by the disorienting influence of the thermal motion of the atoms. As we increase the field, more and more moments will orient themselves along the field, but if we increase the temperature this orientation will be opposed. At high enough field at a temperature T all moments would line up along the direction of the field. This will give us the maximum possible magnetic moment $N\vec{m}$. At low

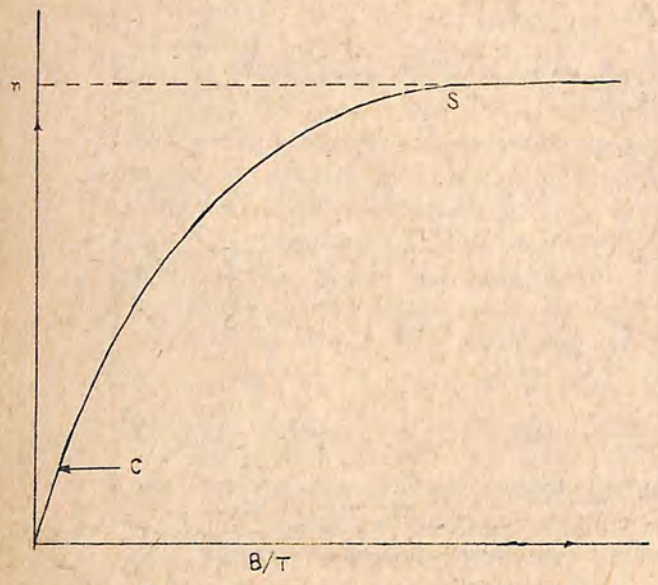


Fig. 17.18 Total magnetic moment (M) of a paramagenetic solid as a function of $\frac{B}{T}$; S—Saturation region ; C—Curie's law region.

fields the total magnetic moment would be directly proportional to the magnetic field $\vec{B}$ and inversely proportional to temperature T. As one increases the field or lowers the temperature the total moment would reach the saturation value $\vec{Nm}$. Fig. 17.18 shows how total magnetic moment varies as a function of $\frac{B}{T}$. At low $\frac{B}{T}$ when we are far from the saturation region we can see that the magnetisation M (which is defined as the total magnetic moment per unit volume) would be given by

$$M = C\frac{B}{T}$$

Where C is a constant. This relation is known as Curie's law. This was discovered by Pierre Curie experimentally.

We thus see that for a solid to be paramagnetic it is necessary that some of its constituent atoms or ions should possess a magnetic moment M. If they do not possess a moment then there will be no paramagnetic effect and the material would be diamagnetic. As has been said before the presence or absence of an atomic moment is decided by the electronic structure of the atom or the ion.

Ferromagnetic Materials

The magnetisation observed in ferromagnetic materials is much larger than in paramagnetic substances. The elements iron, cobalt, nickel, gadolinium and dysprosium and a number of alloys are ferromagnetic. In a ferromagnetic material the magnetic moments of the individual atoms are coupled to neighbouring moments very strongly through an interaction which has come to be known as exchange interaction. The exchange interaction in ferromagnetic materials align the moments all in the same direction as sketched in Fig. 17.19.

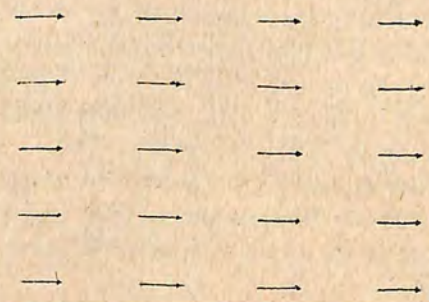


Fig. 17.19 In a ferromagnetic material the exchange interaction aligns all the moments along the same direction.

In a ferromagnet if all the moments are lined up parallel to each other why is it that every piece of iron or nickel is not a magnet? The reason is that an ordinary piece of unmagnetized ferromagnet consists of numerous domains in each of which the moments

are pointing along the same direction. But in neighbouring domains the moments of a domain are pointing in different directions. A piece of a ferromagnet consists of a very large number of domains all pointing in different directions and no direction is preferred over others. This means that the total moments due to domains will cancel each other and the piece of the ferromagnet will appear unmagnetized. Fig. 17.20 shows the orientations of domains in a ferromagnetic material. In the presence of a magnetic field the domains align themselves along the direction of the field to give a strong magnetization.

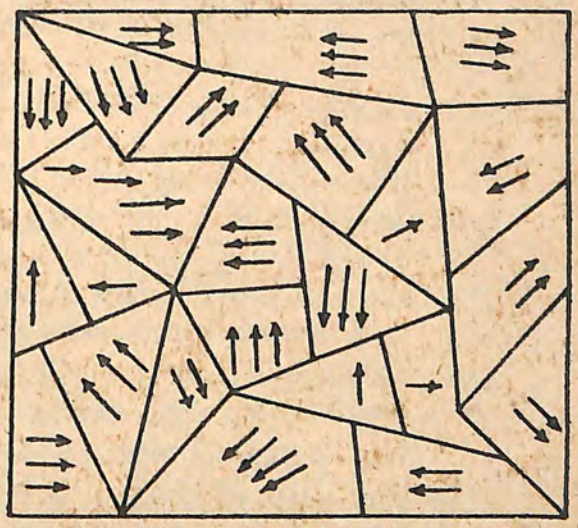


Fig. 17.20 The orientation of domains is such that a piece of ferromagnetic material appears unmagnetised.

Exercises

17.1 A 4 m long aluminium wire whose diameter is 3 mm is used to support a mass of 50 kg. What will be the elongation of the wire. Y for aluminium is 7×10^{10} N/m² (3.96 mm)

17.2 A cable is replaced by another one of the same length and material but of twice the diameter. How will this affect the elongation under a given load ? How does this affect the maximum load it can support without exceeding the elastic limit ?

17.3 A wire increases by 10^{-3} of its length when a stress of 1×10^{8} N/m² is applied to it. What is the Young's modulus of the material of the wire ?
(1×10^{10} N/m²)

17.4 Graphite consists of planes of carbon atoms. Between atoms in the planes there are strong interatomic forces. Between the atoms in different planes there are only very weak forces. What kind of elastic properties you expect for graphite ?

17.5 Consider a steel bridge 200 m long in a locality where temperature varies from 243 K to 313 K. Find the change in the length of the bridge for the above seasonal variation in the temperature ?
Coefficient of linear expansion of steel 11×10^{-6}/ K. (Ans : 15.4 cm)

17.6 A circular hole of radius 1cm is drilled in a brass sheet kept at 293 K. What will be the diameter of this hole when the sheet is heated to 393 K.
(Ans : 2 cm)

17.7 We have a hollow sphere and a solid sphere of equal radii and of the same material. They are heated to raise their temperatures by equal amounts. How will the change in their volumes due to volume expansion be related ?

17.8 Outdoors in winter why does a piece of metal feel colder than a piece of wood ?

17.9 What would be the effect of temperature on the dielectric constant of a substance containing molecules with permanent dipole moment ?

17.10 What kind of thermal conductivity and specific heat requirements would you specify for a cooking utensil ?

17 11 Why do electrons in insulators not contribute to heat conductivity ?

17.12 Why should alkali halides be diamagnetic rather than paramagnetic ?

UNIT 18

Nuclear Physics

18.1 Atomic Nucleus

In unit 9 we have studied the physics of the atom. In Bohr model of the hydrogen atom the hydrogen nucleus is identified as a single positively charged particle called proton. It has a mass of 1.6726×10^{-27} kg. which is about 1836 times the mass of an electron. Its charge is the same as the charge on an electron, viz., 1.60×10^{-19} C but is positive. The mass of the total number of electrons in any atom is negligibly small compared with the mass of its nucleus. Therefore, the main contribution to the atomic mass is from the nucleus only. From the results of Rutherford's scattering experiments (Section 9.3) nuclear size is estimated to be of the order of 10^{-14}m whereas the diameter of an atom is of the order of 10^{-10}m. Hence most of the atom is empty. In 1932 James Chadwick discovered neutral particles called neutrons. The mass of the neutron is 1.6749×10^{-27} kg which is slightly greater than that of the proton and it has no electrical charge. For instance, oxygen nucleus having a charge of + 8e and mass 16 must contain 8 protons. The rest of the mass is accounted for by assuming the presence of 8 neutron's in the nucleus. The constitution of nuclei more complex than hydrogen became clear after the discovery of the neutron. Helium nucleus has a charge + 2e and mass 4 units. Therefore helium nuclei contain two protons and two neutrons each. Like-wise Lithium nucleus which has charge + 3e and mass 7 units contains three protons and four neutrons.

18.2 Isotopes

Protons and neutrons which are present in the nuclei of atoms are referred to collectively as nucleons. The number of protons in a nucleus is called its atomic number Z. The total number of nucleons in the nucleus is generally denoted by A. Hence A — Z gives the number N of neutrons in a nucleus. The chemical properties of elements are determined entirely by the electronic configuration of their atoms. We often come across atoms whose electronic structures are identical but whose nuclear masses differ. They are called isotopes. For example, Deuterium nucleus (Deuteron) has one proton and one neutron and Tritium nucleus (Triton) contains one proton and two neutrons. Since Deuterium and Tritium have identical electronic structure like hydrogen they are called *Isotopes of hydrogen.*

Since the neutron has nearly the same mass as the proton, the number A, which

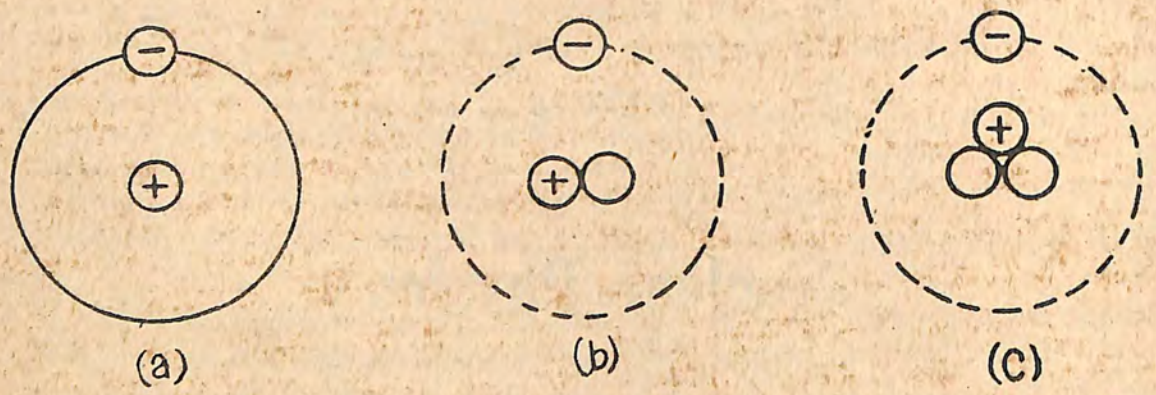


Fig. 18.1 Isotopes of Hydrogen.

represents the sum total of the protons and neutrons in the nucleus is referred to as the mass number of the isotope. The conventional way of representing any isotope is by $^{A}_{Z}x$ where x is the chemical symbol of the element, Z its atomic number and A the mass number of the isotope. According to this notation hydrogen is designated as $_1H^1$ and its isotopes deuterium and tritium $_1H^2$ and $_1H^3$ respectively (Fig. 18.1). Lithium has two stable isotopes $_3Li^6$ and $_3Li^7$ (Fig. 18.2) and the isotopes of chlorine are $_{17}Cl^{35}$ and $_{17}Cl^{37}$. Normal chlorine is a mixture of these two isotopes and has an atomic weight 35.45.

When an atom is talked of with particular reference to its nuclear composition it is called a *nuclide*. So istopes are nuclides having the same value of Z. Those having the same value of the atomic mass number A but different values of Z are called *Isobars*.

18.3 Mass Spectrometer

A mass spectrometer is an instrument for measuring the atomic masses. Fig. 18.3.

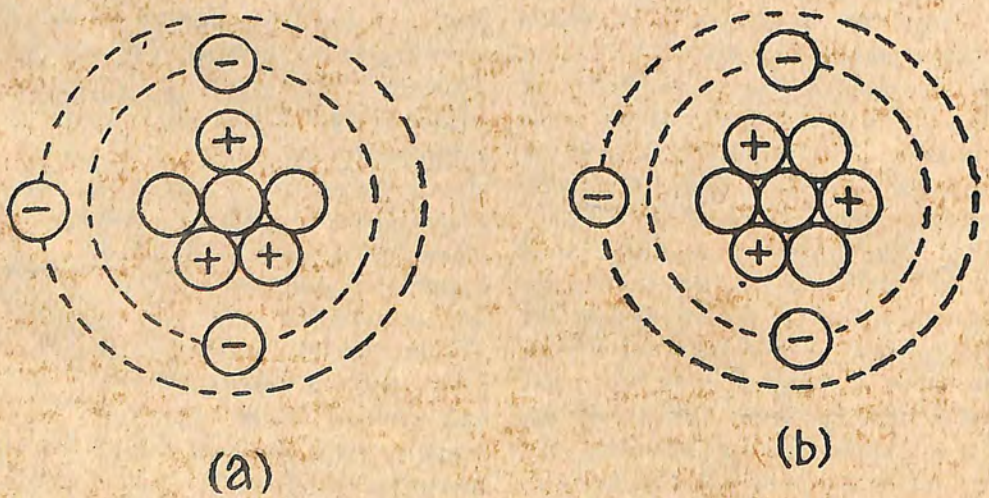


Fig. 18.2 Isotopes of Lithium

indicates the arrangement in one type of mass spectrometer. A beam of positive ions from a source is allowed to pass through a long tube T with slits S_1 and S_2 at each end. As they enter through slit S_1 into the tube T, the ions are subjected to electric field E. These ions are also simultaneously subjected to a magnetic field B^1 which is at right angles to the electric field. The sense of the electric field is such that the ions are deflected from left to right whereas the sense

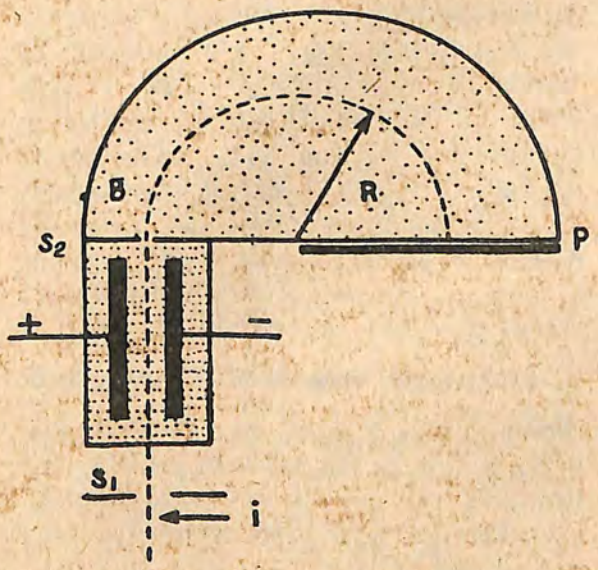


Fig. 18.3 A mass spectrometer.

of the magnetic field B^1 (into the plane of paper) is such that it deflects the particles from right to left. Thus the ions experience two opposing forces of magnitude Ee and B'eV. As the slit S_2 is exactly opposite to S_1 only those undeflected ions will pass through S_2 into the semi circular chamber of the spectrograph. This condition is satisfied only by those ions having a velocity V which satisfies the condition:

$$B'eV = Ee \text{ or } V = \frac{E}{B'} \qquad \ldots (18.1)$$

In the region corresponding to the semi-circle in Fig. 18.3 there is only uniform magnetic field B perpendicular to the path of the charged particle. Hence it moves in a circular path and the magnetic force on the particle provides the necessary centripetal force. The radius R of the path of the particle is given by

$$\frac{mV^2}{R} = BeV$$

$$\text{and } R = \frac{m}{e}\frac{V}{B} \qquad \ldots (18.2)$$

Substituting the value of V from 18.1 in 18.2 we get

$$\frac{e}{m} = \frac{E}{BB'} \cdot \frac{1}{R} \qquad \ldots (18.3)$$

R is determined by keeping a photographic plate as shown in Fig. 18.3. From a knowledge of the values of E, B and B' and measuring R, the value of $\frac{e}{m}$ can be obtained from equation (18.3). Using the known value of unit charge e the atomic mass m can be calculated.

The discovery of isotopes has cleared the apparent anomaly of fractional atomic mass numbers.

18.4 Nuclear Mass

From scattering experiments the radii of the nuclei of oxygen and lead were found to be 3×10^{-15}m and 7×10^{-15}m respectively. The masses of these nuclei are 2.7×10^{-26} kg. and 3.4×10^{-25}kg. Let us calculate the densities of their nuclei.

The volume of the nucleus can be found making use of the formula $\frac{4}{3}\pi r^3$ for volume of a sphere of radius r. And density is given

by $\frac{\text{mass}}{\text{volume}}$.

Density of oxygen nucleus

$$= \frac{3 \times 2.7 \times 10^{-26}}{4\pi(3 \times 10^{-15})^3}$$

$$= 2.4 \times 10^{17} \text{kg/m}^3$$

Density of lead nucleus

$$= \frac{3 \times 3.4 \times 10^{-25}}{4\pi(7 \times 10^{-15})^3}$$

$$= 2.4 \times 10^{17} \text{kg/m}^3$$

We find that the density is the same for both nuclei. The value of this density is very high.

According to the liquid drop model of an atomic nucleus different nuclei are considered as minute droplets of nuclear matter. The fact that the density is the same in the case of both the nuclei in the above example is in accordance with the nuclear theory based on *liquid drop model.* The mass of one cubic centimeter of nuclear matter packed together will be 240 million metric tons.

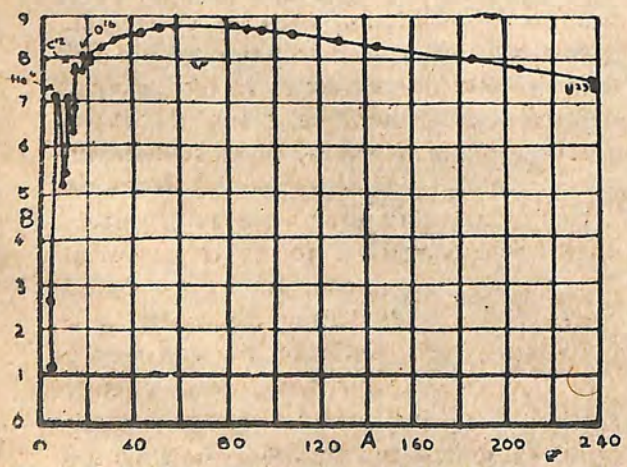


Fig. 18.4 Binding energy per nucleon plotted as a function of mass number A.

18.5 Electron—Volt and Atomic Mass Unit

An electron—volt (ev) is the amount of energy that would be imparted to an electron (charge 1.602×10^{-19}C) on being accelerated through a potential difference of 1 volt.

Hence 1 ev $= 1.602 \times 10^{-19} \times 1$ J

$= 1.602 \times 10^{-19}$ J

The size of this unit is convenient for measuring energies in atomic reactions. For example, to remove the one electron from an atom of hydrogen 13.6 ev of energy is needed.

Nuclear forces are very much larger than atomic forces. Therefore, while dealing with nuclei, a much bigger unit, a million — electron—volt is used. It is denoted by Mev and is equal to 10^6 ev. For example, about 2.2 Mev energy will be needed to separate the neutron from the proton in the nucleus of heavy hydrogen (Deuterium).

At present atomic masses are expressed in terms of the actual mass of ${}_6C^{12}$ isotope of carbon. The unit of atomic mass or the atomic mass unit (abbreviated as amu) is $\frac{1}{12}$ of the actual mass of the ${}_6C^{12}$ isotope of carbon. This is equal to 1.66×10^{-27}kg which is close to the mass of an atom of hydrogen. We can now calculate the energy equivalent of one amu using Einstein's mass-energy equation

$$E_o = m_o c^2$$

Where E_o is the energy equivalent to a rest mass m_o and c is the speed of light. From the above equation

1 amu $= (1.66 \times 10^{-27})(2.998 \times 10^8)^2$ J

on dividing this by 1.602×10^{-19} J, we get energy in electron volts as equal to 931×10^6 ev or 931 Mev.

$$\boxed{1 \text{ amu} = 931 \text{ Mev.}}$$

18.6 Binding Energy

Deuteron is the deuterium nucleus and it consists of only one proton and one neutron.

Neutron Mass = 1.008665 amu
Hydrogen Mass = 1.007825 amu
Total Mass = 2.016490 amu

But the atomic mass of deuterium is found to be 2.014103 amu which is less than the mass of the particles put together. The difference is 0.002387 amu. In energy units it is equal to 2.22 Mev. This is accounted for as the energy that binds the two nucleons together. The surplus energy which nucleons give up by virtue of their mutual attractions when they become bound together to form a nucleus is called the *binding energy*. When a neutron and proton come together a small part of their mass is radiated in the form of electro-magnetic radiation. The two nucleons attract each other so strongly that, it has been found experimentally, that only γ rays of energy 2.22 Mev or greater can pull the proton and neutron apart.

In the above example of deuteron the binding energy is 2.22 Mev. The total binding energy of Neon $_{10}Ne^{20}$ nucleus is 160.6 Mev and that of Bismuth $_{83}Bi^{209}$ nucleus is 1640 Mev. If these binding energies are divided by their respective mass numbers 20 and 209 we get 8.0 Mev and 7.8 Mev respectively. They represent the binding energy per nucleon in each case. Even though the two nuclides that are considered lie almost at the extremes of the mass number scale their binding energies per nucleon are approximately equal. In Fig. 18.4 the binding energy per nucleon is plotted against mass number. With the exception of $_2He^4$, $_6C^{12}$ and $_8O^{16}$ the values lie on or in the proximity of a curve. The curve has a broad maximum close to the value 8.5 Mev/nucleon in the mass number range from about 40 to 120 As the mass number increases the binding energy per nucleon gradually decreases and falls to 7.6 Mev/nucleon for uranium. The decrease in binding energy per nucleon for large mass numbers is due to the coulomb repulsion between the protons which makes the nuclei increasingly less stable.

18.7 Nuclear Forces

What is the type of force that is responsible for the binding of nucleons in a nucleus? The infinitely small size of the nucleus in which all the protons and neutroes are closely packed indicates that there must be strong attractive forces to hold nucleons (protons and neutrons) together. These attractive forces cannot be electrostatic forces, for protons, being of like charge would have repelled when held close together and neutrons have no charge. The electrostatic forces between protons being repulsive or disruptive, they cannot account for the strong attractive forces between nucleons which are responsible for the large binding energy in a nucleus of the order of 8 Mev per nucleon. The gravitational force is a force of attraction between every pair of nucleons but it is infinitesimally smaller than the powerful attractive forces between nucleons. It is, therefore, assumed that the purely attractive forces which exist between nucleons are of a new type which have no analogy with other forces such as gravitational and electrostatic forces in classical physics. These attractive forces are called *nuclear forces*. We shall now consider some important characteristics of nuclear forces. The electrostatic coulomb forces operate between any pair of charges at any distance following inverse square law. The situation in the case of the nucleus is entirely different. All the nucleons are closely packed in the tiny nucleus like a set of marbles in a box. The forces which hold the nucleons together must be those which

exist between the individual neighbouring nucleons in the nucleus which are separated by distances of the order of 10^{-15}m. The nuclear forces between nucleons are, therefore, *short range forces* operating over very short distances only of the order of 10^{-15}m.

These short range forces operate only between two neighbouring nucleons and they are of three types. These are — (1) the force between a proton and neutron (pn force) (2) the force between two protons (pp force), and (3) the force between two neutrons (nn force). All these are attractive forces. It may be noted that, in the case of (pp) nuclear force, there is also a repulsive force between the two protons, but this is weak compared to the strong nuclear force. There is enough experimental evidence to show that the (pn), (pp) and (nn) nuclear forces are approximately equal. This is the *charge independence character* of nuclear forces.

18.8 Nuclear Reactions

An important landmark in the progress of nuclear physics was the discovery of artificial transmutation of elements by Rutherford in 1919. Transmutation is transformation of one element into another. He was the first to artificially transmute or change the element nitrogen into oxygen by bombarding nitrogen nuclei with high energy α particles from a radioactive source. Fig. 18.5 shows the schematic diagram of the apparatus used by Rutherford. T is a tube filled with nitrogen gas. α particles from a radio-active source R were allowed to bombard the nitrogen nuclei. A is a thin silver disc covering the opening at the end of the tube and S is a fluorescent screen for observing scintillations through a microscope M.

When α particles bombarded nitrogen nuclei, scintillation, were observed on the screen S due to nuclear particles which were identified as protons. The nuclear reaction that has taken place may be written symbolically as

$$_2He^4 + {}_7N^{14} \rightarrow {}_8O^{17} + {}_1H^1 \quad \ldots \quad (18.4)$$

$_8O^{17}$ is an isotope of oxygen. Starting with two stable nuclei namely, helium and nitrogen, two new stable nuclei oxygen and hydrogen were produced. This process is, therefore, referred to as artificial transmutation of elements.

Conservation Laws in Nuclear Reactions

It is found that in any nuclear reaction there is conservation of (a) Momentum (b) Charge, and (c) Energy.

(a) In any nuclear reaction the total momentum of the Atoms entering into reaction remains the same as the total momentum of the products after the reaction.

(b) In the nuclear reaction stated above

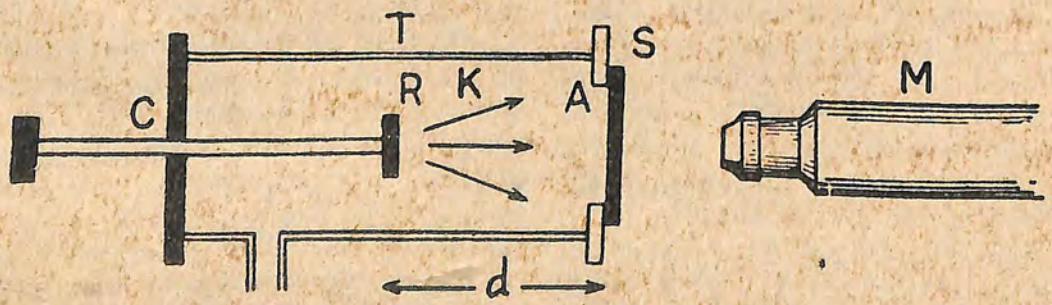


Fig. 18.5 Schematic diagram of Rutherford's apparatus.

the total charge of the items entering into reaction is $2 + 7 = 9$ units and that after the reaction is $8 + 1 = 9$ units. Therefore, the total charge is conserved.

(c) In relativistic mechanics the energy of any particle is the sum total of rest-mass energy and kinetic energy. The combination is effected using Einstein's equation $E_0 = m_0c^2$ which expresses the equivalence of mass and energy. According to the principle of conservation of energy the total energy before a reaction equals the total energy after the reaction. Nuclear energy equations are generally written in atomic mass units. But any consistent units of mass or energy can be used.

The mass energy equation for the nuclear reaction just studied may be written as :

$m_1c^2 + {}^Ek_1 + m_2c^2 + {}^Ek_2 = m_3c^2 + {}^Ek_3 + m_4c^2 + {}^Ek_4$ ... (18.5) Where m_1, m_2, m_3, m_4, are the rest-masses of ${}_2He^4$, N^{14}, ${}_8O^{17}$ and ${}_1H^1$ respectively and ${}^Ek_1$, ${}^Ek_2$, ${}^Ek_3$, and ${}^Ek_4$ their respective kinetic energies.

Note : If, initially, energy is also supplied in the form of a photon and if a photon is emitted finally, then the energies of these photons also should be included in the equation.

Let us express all the quantities in the equation in atomic mass units. The kinetic energy ${}^Ek_2$ of ${}_7N^{14}$ is very small (fraction of one Mev). Hence it can be neglected in comparison with other quantities. The energy of the α particle in this experiment was found to be 7.7 Mev. Expressed in atomic mass units.

$$7.7 \text{ Mev.} = \frac{1}{931} \times 7.7 = 0.008270 \text{ amu}$$

Neglecting ${}^Ek_2$, the masses of the items on the left hand side of equation (18.5) are:

$m_1 = {}_2He^4$	$= 4.002604$	amu
$m_2 = {}_7N^{14}$	$= 14.003074$	amu
${}^Ek_1$	$= 0.008270$	amu
Total	$= 18.013948$	amu

Coming to the right side of the equation

$m_3 = {}_8O^{17}$	$= 16.999133$	amu
$m_4 = {}_1H^1$	$= 1.007825$	amu
$m_3 + m_4$	$= 18.006958$	amu

Substituting the above values in equation 18.5 we get

$${}^Ek_3 + {}^Ek_4 = 18.013948 - 18.006958$$
$$= 0.006990 \text{ amu}$$
$$= 931 \times 0.006990 = 6.508 \text{ Mev.}$$

6.508 Mev. energy is shared as kinetic energy by the proton and oxygen nucleus.

Chadwick in 1932 conducted an experiment in which he allowed α particles (D) from polonium to strike beryllium metal (B) (Fig. 18.6). Penetrating rays were found to come out of the beryllium metal. They were found to be particles without charge. They are called neutrons (N). The reaction may be written as

${}_2He^4 + {}_4Be^9 \longrightarrow {}_0n^1 + {}_6C^{12}$ ... (18.6)

${}_0n^1$ denotes the neutron since its charge is zero and mass nearly the same as that of a proton.

Instead of α particles high speed protons also can be used as projectiles to study the disintegration of nuclei. In one experiment

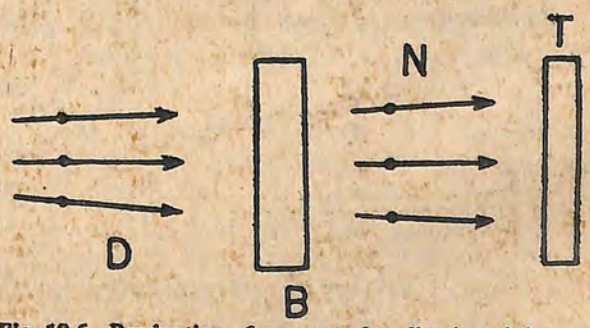


Fig. 18.6 Production of neutrons by allowing alpha particles to strike beryllium metal.

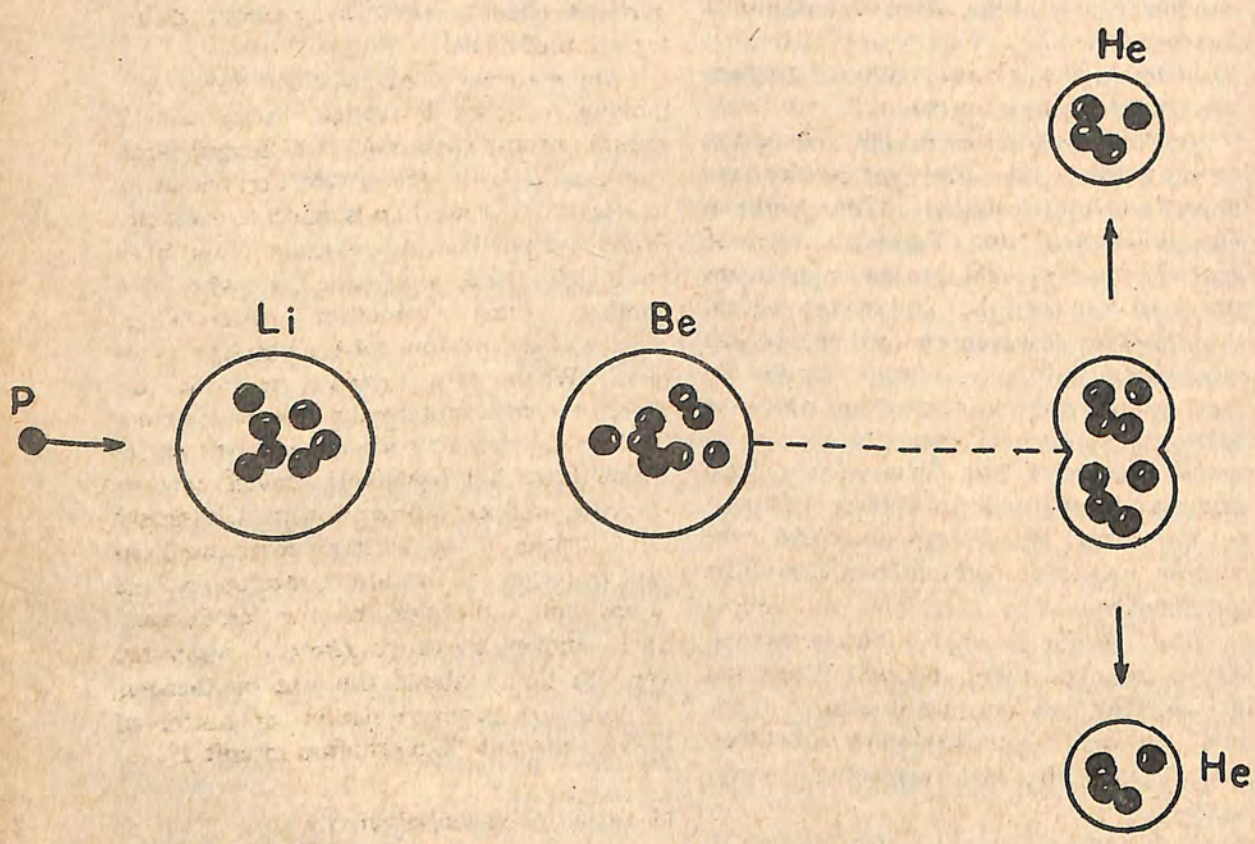


Fig. 18.7 Disintegration of a lithium nucleus by a proton.

protons accelerated to energies of the order of 0.15 Mev were used to bombard Lithium metal. When a proton collides with a lithium nucleus and is captured, two α particles are produced (Fig. 18.7). Neglecting E_{k_2} the reaction may be writtens as

$$_1H^1 + {}_3Li^7 + E_{k1} \longrightarrow 2\,(_2He^4) + 17 \text{ Mev.} \quad \ldots (18.7)$$

The energy of each α particle was measured to be about 8.5 Mev.

18.9 Properties of Neutrons

We have seen that a neutron is a particle with no charge and with a mass slightly higher than that of the proton. Neutrons are present in the nuclei of all elements except hydrogen. Although in a nucleus they appear to last for ever neutrons outside a nucleus have an average life of about 1000 seconds before they decay. A free neutron spontaneously decays into a proton. electron and an antineutrino.*

$$_0n^1 \longrightarrow {}_1H^1 + {}_{-1}e^0 + \overline{\nu^0}$$

The electron charge being -1 the subscript to e is -1. The mass of an electron being about $\frac{1}{1836}$ proton, it is negligibly small compared to the mass of a

* antineutrino is an elementary particle (See-Unit. 19)

proton. Hence the mass of electron is written as 0.

Being without charge neutrons are not subject to Coulomb forces and can easily penetrate even heavy nuclei. Therefore, neutrons serve as ideal projectiles for inducing nuclear reactions in the study of the properties of nuclei. In the nuclei of heavier elements there are more neutrons than protons for the following reason. Protons being positively charged repel each other and in order to maintain stability of the nucleus more neutrons become necessary in the case of heavier elements.

By keeping a thin plate of beryillum metal in the path of an intense beam of deuterons from a cyclotron (an instrument used for accelerating atomic particles) high speed neutrons are produced. Substances of known chemical composition are inserted in the path of these neutrons and the disintegration products are studied.

Following are some examples of neutron reaction using high-speed neutrons :

$${}_0n^1 + {}_9F^{19} \longrightarrow {}_7N^{16} + {}_2He^4$$

$${}_0n^1 + {}_{27}Co^{59} \longrightarrow {}_{27}Co^{60} + \gamma \text{ ray}$$

$${}_0n^1 + {}_{80}Hg^{198} \longrightarrow {}_{79}Au^{198} + {}_1H^1$$

The last reaction indicates the transmutation of mercury into gold. Thus the alchemists' dream of producing gold from mercury has been realised. However, this method cannot be adopted, in practice, for obtaining gold from mercury because it would be very expensive.

Thermal Neutrons : Since a neutron does not experience a repulsive force as it approaches a nucleus the chances of its penetrating the nucleus are more. Hence disintegration of a nucleus by slow moving neutrons is possible while low speed charged particles like protons, deuterons or α particles cannot easily bring about disintegration of nuclei.

Any material used to slow down fast moving neutrons is called "moderator". Paraffin or any such material having large number of hydrogen atoms serves as a moderator. Deuterium is another example. When two particles of the same mass have an elastic head on collision with one another their velocities interchange. Neutrons and protons have nearly the same mass. When fast moving neutrons are passed through paraffin the neutrons make elastic collisions with hydrogen nuclei which have comparatively much smaller velocities and are slowed down. Ultimately the veloeities of the neutrons correspond to the velocities of random motion of the atoms and molecules of the moderator. Such neutrons are called *Thermal Neutrons.* We will study about the use of thermal neutrons to produce fission of nuclei of heavier elements like uranium in unit 19.

18.10 Artificial Radio activity

The phenomenon of artificial or induced radio activity was discovered by F. Joliot and Irene Curie Joliot. They exposed various substances to α rays from naturally radioactive elements and made extensive study of the disintegration produced in them.

In one experiment α rays from polonium were allowed to strike an aluminium plate (Fig. 18.8) (a). At the end of the experiment, even after the radio-active source was taken away the radiation detector was found to respond to some kind of penetrating radiation. They investigated the nature of these radiations and found them to be positrons* coming from aluminium which

* Positron is a particle having the same mass as an electron but with equal positive charge (See Unit. 19)

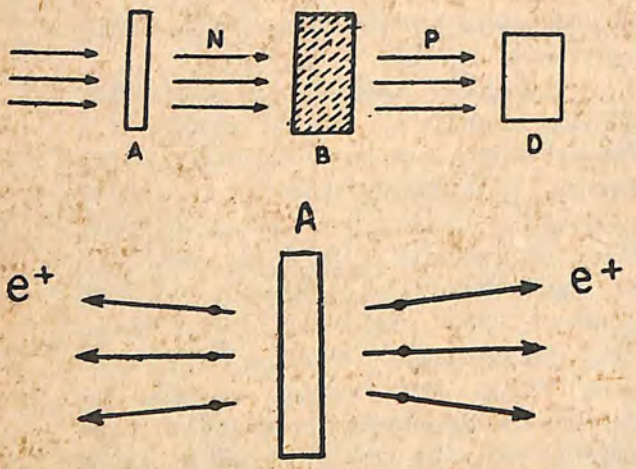


Fig. 18.8 Schematic diagram of the experimental arrangement used by the Curie-Joliots in their discovery of induced radioactivity. N—Neutrons, A—Aluminium, P—Protons, B—Paraffin block, D—Detector.

was formerly exposed to α rays (Fig. 18.8). Further, the emission of positrons decreased with time according to an exponential law. Hence the phenomenon was just like natural radio-activity.

When α rays strike aluminium, neutrons are ejected out and phosphorus $_{15}P^{30}$ is formed according to the reaction

$$_2He^4 + {}_{13}Al^{27} \longrightarrow {}_{15}P^{30} + {}_0n^1 \quad ...(18.8)$$

It is this isotope $_{15}P^{30}$ of phosphorus which exhbited radio-activity. Its half-life is 2.55 minutes. It decays into silicon emitting a positron accompanied by nutrino**

$$_{15}P^{30} \longrightarrow {}_{14}Si^{30} + {}_1e^0 + \upsilon$$

18.11 Radio-Isotopes

In the previous section we have seen that radio-active phosphorus $_{15}P^{30}$ was obtained when aluminium was bombarded by α particles. A variety of radio-active isotopes are produced by adopting similar procedure using besides α particles, deuterons, neutrons, etc,. as projectiles to hit several elements as targets.

Although produced in very small quantity, it is possible to separate the radio-active phosphorus from the aluminium (reaction studies in 18.10) which is non-radio-active. Also it is possible to trace this small amount of radio-active phosphorus through successive chemical processes by using suitable detectors. This fact is made use of in radioactive "tracer techniques". In nuclear reactors, described in unit 19, large number of radio-active isotopes can be produced. Some radio-isotopes have great application as tracers in various fields such as medicine, agriculture and industry. Besides their use as tracers, radio-isotopes have a variety of other applications. In the next section some examples of the use of radio-active isotopes are given.

18.12 Use of Radio-isotopes

(a) *In Medicine*

In medical diagnosis radio isotopes are particularly useful in cases where conventional diagnostic procedures are not quite helpful.

When Iodine is taken into the body, it is absorbed more strongly by the thyroid gland than by other tissues. The radio isotope $_{53}I^{131}$ has a half-life of 8 days and it decays by β-emission. Radioiodine is administered to a patient and by means of sensitive detectors of β rays, information about the size, activity and functioning of the thyroid gland is obtained.

**Neutrino is an elementary particle which carries no charge and has no rest-mass (See unit. 19)

Diagnosis of brain tumors is difficult because of the inaccessibility. Radioiodine and some other isotopes are found useful in the diagnosis of brain tumors. A small quantity of the radioisotope is incorporated in organic dyes which are absorbed strongly by the tumor tissue than by normal tissues. This is injected into the body. Then by means of detectors of the emitted radiation the radiologist gets information about the size and location of the tumor.

Suppose the circulation of blood in a person is restricted. It is easy to locate the position of the constriction by using Radiosodium. Sodium chloride prepared by using radiosodium is taken and a solution of it of proper strength is injected into a vein of the patient. A radiation counter is kept in contact with, say, one of the feet of the patient. If the blood circulation is normal the radio-activity increases rapidly and reaches a maximum value in less than one hour. If the radioactivity increases slowly it means there is difficulty for the blood to reach the foot. By shifting the detector to different parts of the body the exact spot where the constriction exists can be located and necessary treatment provided.

Radiation has long been used in medical therapy for controlling the development and growth of cells, for example, in the treatment of some forms of cancer. In certain cases small capsules or needles containing radium (or its decay products) are implanted within the diseased organ which is thus subjected to the action of gamma rays. At the present time artificial radioisotopes provide more convenient and cheaper alternative sources of radiation both in teletherapy and for implanting in the body. The penetrating rays from the radio-isotope of cobalt Co^{60} have been found to permit higher doses to tissue deep in the body with a minimum of skin reaction.

Leukemia disease is found to respond to treatment by radiation from radio-isotopes of phosphorus. Radiation from radioisotope of iodine has been found to be effective in the treatment of *hyperthyroidism*. It may be noted here that radioisotopes have been used for internal irradiation by taking advantage of the preferred absorption of certain elements in particular organs or tisssues of the body.

(b) In Agriculture

Phosphorus is applied to soil to improve plant growth. Various types of phosphate fertilizers are available and it is helpful to know which variety is utilized most effectively by the particular soil. By using a fertilizer in which radiophosphorus is employed it is possible to know to what extent the plant absorbs phosphorus from that particular fertilizer. This facilitates the selection of the fertilizer most suitable to the soil. Without the use of radio phosphorus it is not possible to distinguish between the phosphorus from the soil and that from the fertilizer.

Crop Mutations: Plant genetic studies have helped in obtaining crops with high yield, resistance to disease and adaptability to new environment. Desirable mutations are induced by irradiation of seeds by γ rays or neutrons. High quality wheat and rice mutants resistant to certain diseases and several jute mutants have been evolved by this process.

Certain insect pests can be controlled by adopting the so-called sterile male technique. The procedure consists of laboratory breeding of large numbers of male insects, sterilizing them with α radiation and releasing them for mating in the infested area. This will bring about rapid reduction in the population of the insect. In

the U.S.A. the screw worm fly, a pest which caused great damage to livestock, was eliminated by adopting the above procedure.

(c) In Industry

Radioactive tracers have wide application in industry both in research and process control. When petroleum products are transported through a single long pipeline; radioactive Antimony-124 used as a tracer, is injected at the source between two different oil stocks and a detector kept outside the pipeline at the receiving end of the line announces the arrival of the new stock of oil. In a similar way the velocity of flow and the degree of intermixing of oil stocks can also be found out.

The purpose of lubricants is to minimise the wear of surfaces, for example of gears and bearings in machinery. It is necessary to select a suitable lubricant for this purpose. A small quantity of a radioactive isotope of iron may be incorporated in the steel of a bearing under test. The lubricating oil is drawn and checked periodically for radioactivity caused by the wearing away of the steel bearing. This gives and idea of the effectiveness of the lubricant.

Gamma rays from a Co^{60} source are found to have the effect of a catalyst in the combination of hydrogen bromide and ethylene. Large quantities of ethyl bromide are produced by this process. Polyethylene, which has great commercial importance, is also produced by Cobalt-60 gamma ray bombardment of ethylene. Irradiated polyethylene has high heat and electrical resistance properties and serves as excellent electrical insulating coating for wires.

In industry quality control of castings and inspection of welded pipes and fabricated machine parts can be made by radiography using cameras housing radio-isotopes like iridium-192 and cobalt-60.

Radio-tracer techniques have been developed for the study of movement of silt on the sea bed. These studies are of vital importance especially in planning dredging operations near ports. A major study of this kind undertaken in recent years was at the Sethusamudram canal project in the Palk Straits. In hydrology, radio-tracers have been used for studying seepage in dams.

(d) Radiocarbon Dating

Carbon dating is a recently developed technique for the dating of cultural history, especially for the dating of items of animal or vegetable origin. C^{14} is a radio-isotope of carbon for which there is "radioactive equilibrium" in the atmosphere. The equilibrium is between production of C^{14} by cosmic rays and its diminution by radioactive disintegration. The half-life of this isotope is 5600 years. All animals and plants have a certain amount of C^{14} which remains in equilibrium with the C^{14} in the atmosphere. After the death of an organism the C^{14} in the body keeps on disintegrating as per exponential decay formula without further replenishment of C^{14}. For example, old bones contain less C^{14} than new bones and old wood contains less C^{14} than new wood. By measuring the ratio of concentration of C^{14} to C^{12} in the remnants of the dead organism and comparing it with the normal ratio in a living organism, it is possible to estimate how long ago the organism ceased to live. This method is now widely used by anthropologists and archaeologists and even historians for finding the age of ancient excavations, manuscripts etc. This method which has become a standard technique is known as radioactive dating.

Exercises

(Where necessary the following data may be used :
Mass of ${}_0n^1 = 1.008\ 665$ amu; mass of ${}_1H^1 = 1.007\ 825$ amu ;
mass of ${}_2He^4 = 4.002\ 604$ amu)

18.1 What is 'atomic number' ? What are 'isotopes' ?

18.2 What is the number of protons and the number of neutrons in each of the following nuclei :
${}_3Li^6$; ${}_5B^{10}$; ${}_6C^{12}$; ${}_{10}Ne^{22}$; ${}_{15}P^{31}$; ${}_{38}Sr^{88}$;
${}_{56}Ba^{137}$; ${}_{92}U^{238}$

18.3 The mass of ${}_{17}Cl^{35}$ is 34.9800 amu. Calculate its binding energy.
What is the binding energy per nucleon ?
(288 Mev : 8.2 Mev)

18.4 Calculate the binding energy per nucleon in the nuclei
${}_2He^4$ and ${}_{26}Fe^{56}$ given mass of ${}_{26}Fe^{56} = 55.934\ 932$ amu.
(7.1 Mev ; 8.8 Mev)

18.5 Calculate the binding energy per nucleon for ${}_{15}P^{31}$ given mass of ${}_{15}P^{31} = 30.973\ 763$ amu
(8.5 Mev)

18.6 Calculate the energy released in the reaction :
${}_3Li^6 + {}_0n^1 \rightarrow {}_2He^4 + {}_1H^3$
given mass of ${}_3Li^6 = 6.015\ 126$ amu and that of ${}_1H^3 = 3.016\ 049$ amu.
(4.78 Mev)

18.7 A neutron strikes a ${}_5B^{10}$ nucleus with the subsequent emission of an alpha particle. What is the atomic number, mass number and chemical name of the remaining nucleus ?

18.8 Describe briefly Rutherford's experiment in which he discovered artificial transmutation of elements ?

18.9 When neutrons bombard ${}_{16}S^{32}$ nuclei, radioactive phosphorus ${}_{15}P^{32}$ is produced. Write down the reaction equation.

18.10 When ${}_{14}Si^{28}$ nuclei are bombarded by neutrons, protons issue out from the target. Write down the reaction equation.

18.11 Complete the following nuclear reaction equations :
(i) ${}_{13}Al^{27} + {}_0n^1 \rightarrow {}_2He^4 +$
(ii) ${}_{15}P^{31} + \tau \rightarrow {}_0n^1 +$
(iii) ${}_9F^{19} + {}_1p^1 \rightarrow {}_8O^{16} +$
(iv) ${}_6C^{14} \rightarrow {}_{-1}B^0 +$

18.12 What are 'thermal neutrons' ? Explain the action of a moderator in slowing down fast moving neutrons.

18.13 Write a note on the use of radio-isotopes in
(a) medicine, (b) agriculture and (c) Industry.

UNIT 19

Nuclear Interactions

19.1 Fission

In 1939 Otto Hahn and his collaborators discovered that when a uranium nucleus, bombarded with neutrons, captures a neutron it splits into two nearly equal fragments. Since this process resembles fission of cells in biology, the phenomenon is called nuclear fission.

Fig. 19.1 shows the splitting of a

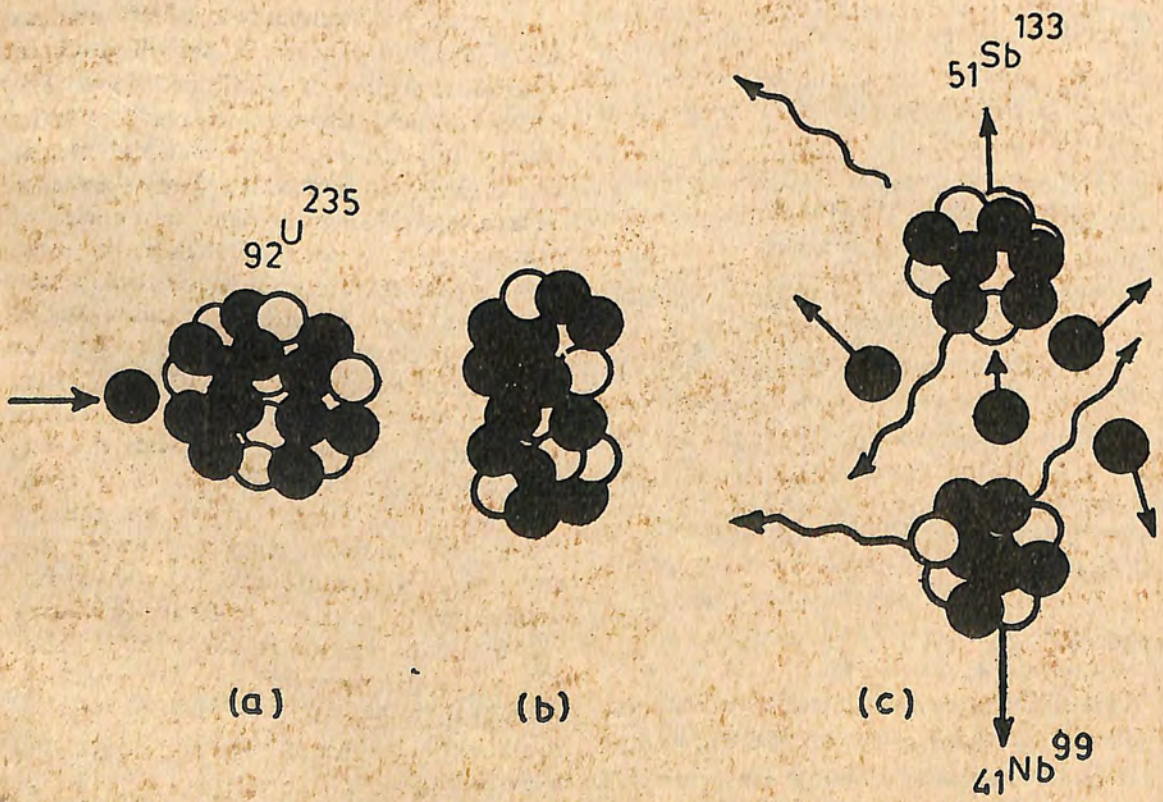


Fig. 19.1 Fission of a nucleus of Uranium $_{92}U^{235}$

nucleus of $_{92}U^{235}$ after it is bombarded by a neutron. The resulting fragments were identified as the nuclides of $_{51}Sb^{133}$ and $_{41}Nb^{99}$. Both these nuclides are unstable and by radioactive disintegration reduce to the stable nuclides $_{55}Cs^{133}$ and $_{44}Ru^{99}$ respectively. Fission is accompanied by the release of neutrons and τ rays.

Uranium found in nature has three principal isotopes. They are as follows:

Isotope	Relative abundance (Per cent)
$_{92}U^{238}$	99.280
$_{92}U^{235}$	0.714
$_{92}U^{234}$	0.006

All the three nuclides are radioactive and they decay by alpha emission. Energy liberation during the fission of uranium is due largely to the isotope U^{235} when slow neutrons are used.

19.2 Energy released in fission

The fission fragments and the accompanying neutrons are released with high velocities. So, large kinetic energies are associated with them. Also energy is released in the form of τ rays. Let us make an estimate of the average energy liberated in the fission of U^{235} nucleus. Let the two fragments have mass numbers say, 100 and 133; then the number of neutrons released is three.

Before fission :

Mass of U^{235}	235.043933 amu
Mass of one neutron ...	1.008665 amu
Total	236.052598 amu

After fission :

Total mass of the two fragments.	232.812000 amu
Mass of 3 neutrons ...	3.025995 ,,
Total	235.837995 ,,

Hence the mass converted into energy is 0.214603 amu, its equivalent in energy units is 200 Mev. nearly. This energy is accounted for as the energy of the gamma rays and the kinetic energy of the fission fragments and the released neutrons.

19.3 Chain Reaction and Controlled Chain Reaction

If a neutron bombards a $_{92}U^{235}$ nucleus and produces fission it will set off a chain reaction in a lump of $_{92}U^{235}$ as follows. The nucleus which captures a neutron splits apart with great violence releasing one or more additional neutrons. These neutrons in turn would be captured by the nuclei of nearby uranium atoms and they, in turn, would split, liberating fresh neutrons which again would be absorbed by uranium nuclei resulting in further fission reactions. Because of the chain reaction, in a very short time, the number of nuclei undergoing fission multiplies in geometric progression. Fig. 19.2 shows a schematic diagram of a chain reaction. It is found that on an average about 10^{-8} second elapses between the emission of a neutron and its subsequent capture by a uranium nucleus. Assuming that the chain reaction is uncontrolled, that is, if all the neutrons that are liberated are allowed to induce further fission, then in about one microsecond (one-millionth of a second), there will be a release of energy of the order of 2×10^{13} Joules, which is a very large quantity. That is why an uncon

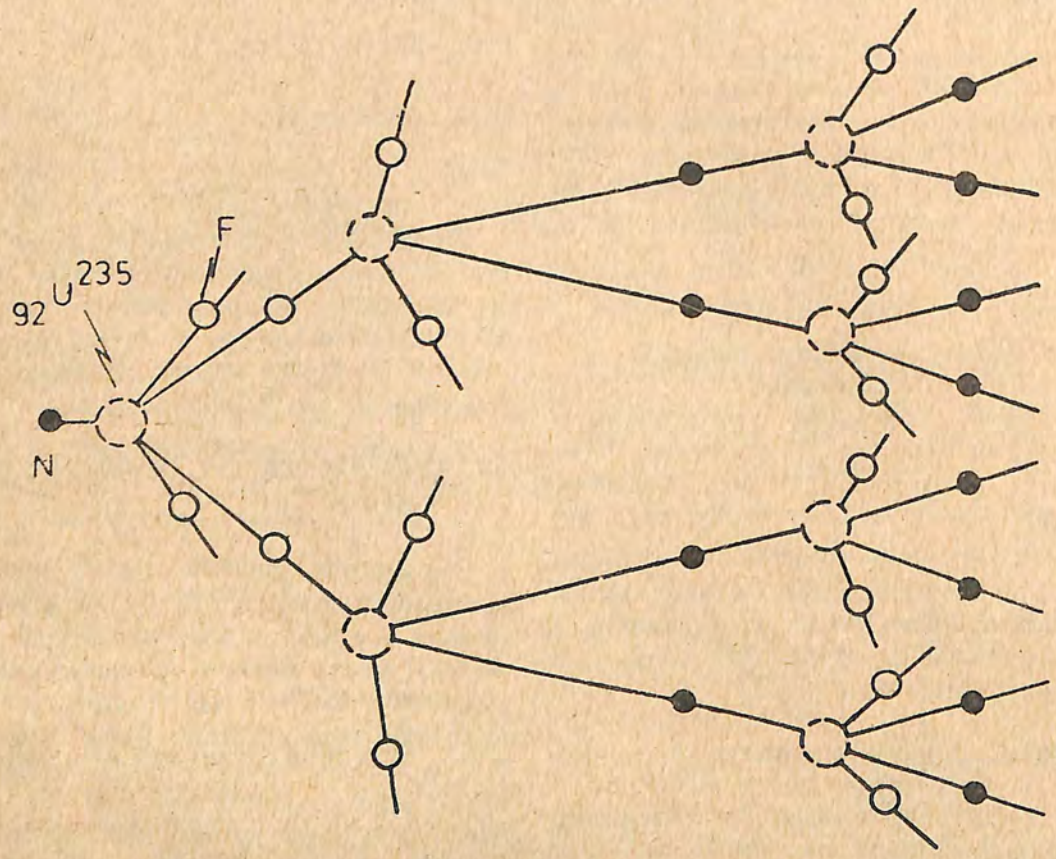


Fig. 19.2 Schematic diagram of a chain reaction in $_{92}U^{235}$.

trolled chain reaction can cause a violent explosion. For this to occur the lump of uranium has to be greater than a certain minimum size.

If the mass of uranium is too small then the liberated neutrons have large scope to escape from the surface and the growth of chain reaction may end before it becomes large enough for an explosion. Hence, there is possibility for an explosion to take place by chain reaction, only if the mass of uranium is greater than some critical size.

In the atomic bomb that was first tested in the U.S.A., $_{92}U^{235}$ was used as the active substance. The requisite quantity of this substance was taken and it was divided into two or more units, each smaller than the critical size. They were kept separated by a short distance. By means of some mechanism they were brought together at the appointed time. Simultaneously neutrons from some source were made to enter the mass and start the chain reaction, finally resulting in a tremendous explosion and release of enormous quantity of energy.

Whether a mass of active material will sustain a chain reaction or not is determined by what is called the *reproduction factors*. If the rate of neutron production is equal to the rate at which neutrons disappear,

the reproduction factor is one and the mass is said to be *critical*.

By the use of moderators the fast neutrons, produced during fission, can be slowed down. And by the use of materials like cadmium, which absorb neutrons without disintegration, the chain reaction can be controlled. Thus, by properly controlling the chain reaction, a constant level of power output can be maintained. Such a controlled chain reaction will be a very efficient source of power. If a conventional power plant needs a consumption of 3 tons of coal per day per 1000 kw, the same output of energy can be produced in a nuclear power plant by using 1g of a suitable isotope as fissionable material per day.

19.4 Nuclear Reactors

A nuclear reactor is a device in which nuclear fission can be maintained as a self-sustained, yet controlled chain reaction. The orignal reactors were known as atomic piles. The first atomic pile was put into operation in Chicago in 1942. By using uranium in the atomic pile, products such as neutrons, radioactive isotopes and heat energy were obtained We have, at present, many kinds of nuclear reactors. The main items in a reactor are :

(1) a quantity of fissionable material,
(2) a moderator for slowing down the neutrons to thermal velocities, and
(3) substances like cadmium for controlling the chain reaction.

The neutrons released by the fission of uranium have energies of the order of 2 Mev. By the use of moderators they can be slowed down to thermal energies of 0.0253 ev (corresponding to velocities of 2200 m/s.) Fig. 19.3 shows the schematic diagram of a

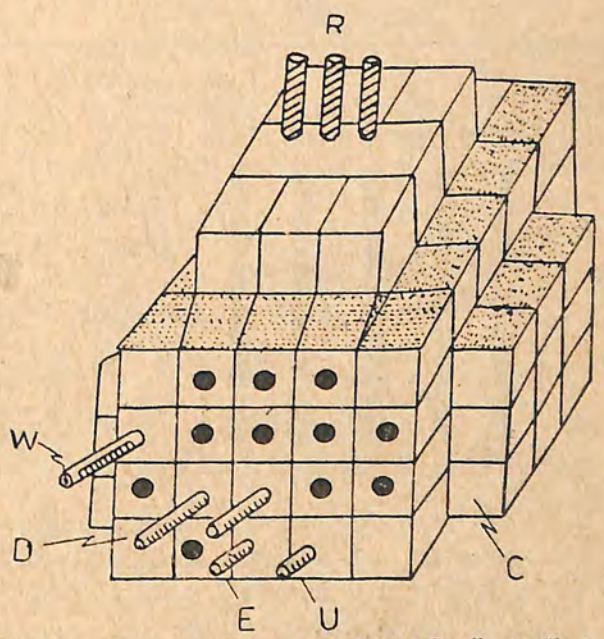


Fig. 19.3 Schematic diagram of an atomic pile (earliest type of nuclear reactor) R—control rods, C—carbon block, W—cooling pipes, D—counting device.

reactor made up of large graphite or carbon blocks used as moderator. There are long cylindrical holes through the blocks for the insertion of fuel elements, control rods, materials to be irradiated and detecting devices. The fuel element is pure uranium metal sealed in aluminum cylinders. Boron or Cadmium is used for control rods, because of their ability to capture slow neutrons. The arrangement of parts in a power reactor shown in Fig. 19.4 helps to understand the working of a reactor.

When a few U^{235} nuclei undergo fission fast neutrons are liberated. As they pass through the surrounding carbon blocks they are slowed down. These thermal neutrons are captured by U^{235} nuclei and fission occurs. By proper arrangement of the control rods (by raising and lowering the rods) the fission events are suitably controlled.

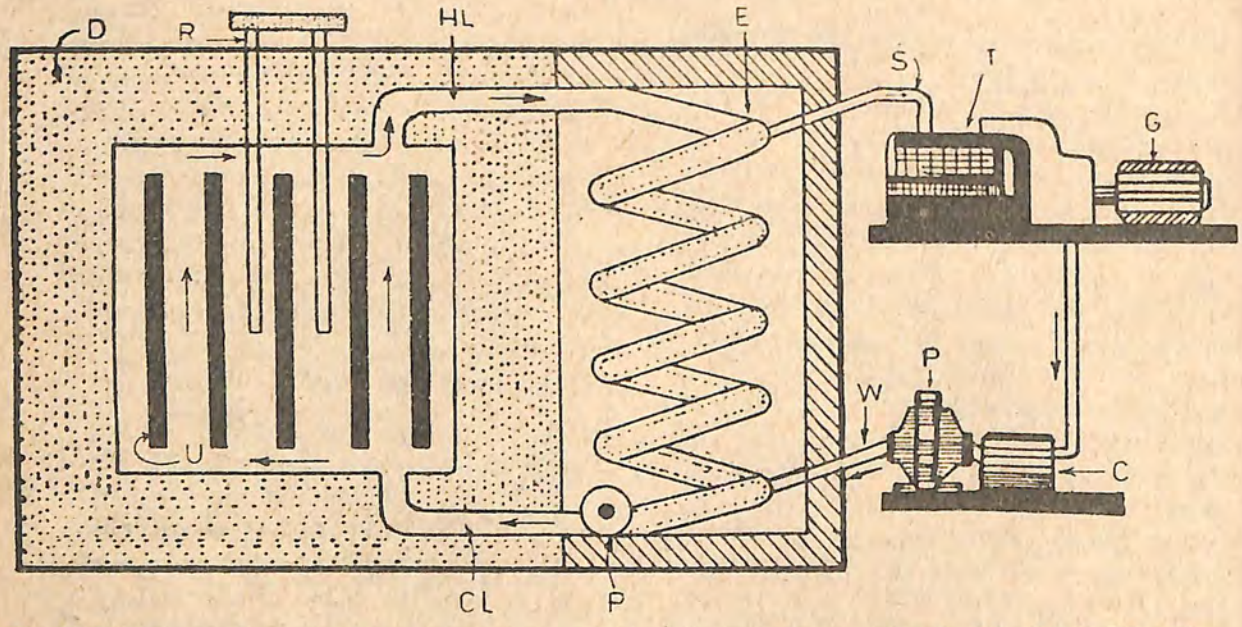


Fig. 19.4 Schematic diagram of a nuclear power plant. P—pump, R—control rods, HL—hot liquid, CL—cool liquid, S—steam, C—condenser.

In a nuclear reactor plutonium ($_{94}pu^{239}$) and many other radio-active isotopes can be produced. Considerable heat is generated in these reactors. This fact is made use of in the construction of nuclear power plants. The heat generated is used for heating suitable coolants which in turn heat water and produce steam. This steam is made to rotate a steam turbine and thereby drive a generator for production of electric power. Fig. 19.4 shows the schematic diagram of a nuclear power plant.

India's First Reactor Apsara

Fig. 19.5 shows a picture of India's first research reactor called Apsara. This was started in 1956 by the Department of Atomic Energy for a variety of needs in the field of nuclear research as well as for the production of radio-isotopes A large number of radio-isotopes are produced by means of this reactor and supplied to various laboratories and hospitals in India for their use.

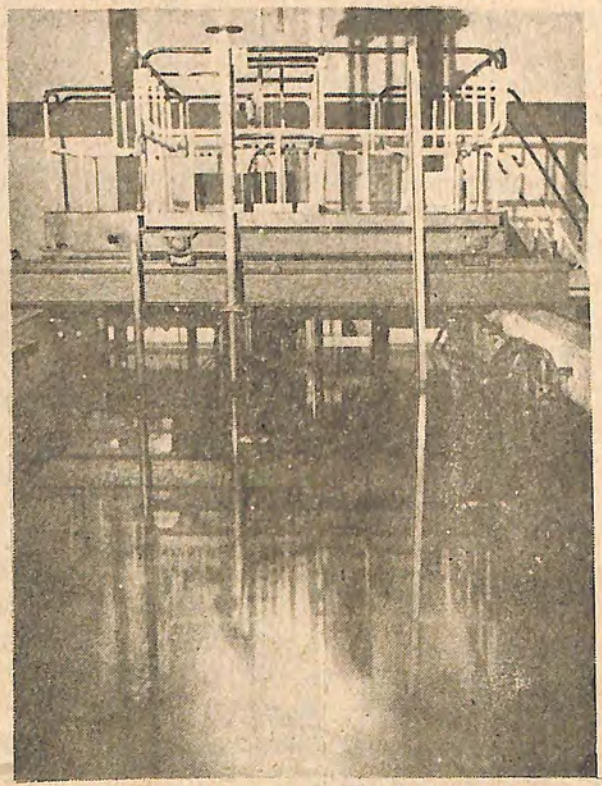


Fig. 19.5 View of Apsara—India's first research reactor at Bhabha Atomic Research Centre, Trombay.

Prior to the commissioning of Apsara in 1956 radiation sources for biological investigations were practically non-existent in India. With the availability of Apsara for research the radiobiological programme in India received a much needed impetus. At the Biology Division of Bhabha Atomic Research Centre, Trombay, neutrons from Apsara are used in genetic studies in agricultural crops. Very useful results have been obtained.

Apsara is a swimming-pool type reactor. A reactor is named according to the moderator or coolant used with it. In the swimming-pool type of reactor a large tank of water is used as a protective shield for the operating personnel. Samples to be irradiated by neutrons are lowered into water and an observer can readily see the blue glow of the water around it (Fig. 19.5). This glow is caused by Cerenkov radiation.*

19.5 Fusion and Energy Generation in Sun and Stars

All animals derive their energy ultimately from plants and plants get their energy from photosynthesis caused by sunlight. Thus all living beings exist because of solar energy. The total radiation emitted by the sun is estimated to be about 3.8×10^{26} Joules of energy per second. The sun has been giving out energy, roughly at this rate, for millions of years. Yet there are no signs of cooling. For a long time this was a mystery to the physicists. But nuclear studies have given a clue to the understanding of this problem.

We have seen that it is possible to produce fission of nuclei of heavy atoms like uranium and plutonium. The reverse of fission, namely, fusion of light nuclei, like protons and deuterons, into helium nuclei is also found possible. Like fission processes, already studied, fusion processes also are accompanied by the liberation of large quantities of energy.

The core of the sun is estimated to be at a temperature of about 20 million degrees C. It is presumed, with the knowledge available at present, that, by a set of nuclear reactions (primarily the proton-proton cycle) taking place at this high temperature enormous amount of energy is liberated. The reactions are :

$$_1H^1 + {}_1H^1 \rightarrow {}_1H^2 + {}_1e^0 + 0.4 \text{ Mev.}$$

$$_1H^1 + {}_1H^2 \rightarrow {}_2He^3 + 5.5 \text{ Mev.}$$

$$_2He^3 + {}_2He^3 \rightarrow {}_2He^4 + 2({}_1H^1) + 12.9 \text{ Mev.}$$

Each of the first two reactions has to occur twice to produce one $_2He^4$ or helium nucleus. The net result is the fusion of four hydrogen nuclei to form one nucleus of helium atom besides producing two positrons and liberating 24.7 Mev. of energy. From the knowledge available at present it is inferred that stars, whose interiors have the same or lower temperature than the interiors of the sun, obtain their energy by the proton-proton cycle, while stars whose interiors are hotter than the core of the sun produce their energy by another cycle of reactions called the carbon-nitrogen cycle. Estimates of hydrogen in the sun suggest that there is enough hydrogen in the sun to keep the proton-proton cycle going for about 30 billion years.

The energy liberated by nuclear fusion is known as *thermo-nuclear energy*. Very high temperatures and densities are needed for fusion to take place. Because of high temperature, the light nuclei have sufficient energy to overcome the forces of electro-

*Fast moving electrons such as β particles from radio-active materials produce light within a transparent medium if their velocity is greater than the velocity of light in that medium. This in known as Cerenkov radiation

static repulsion and come together. Because of high density, collisions take place frequently. These conditions for fusion to take place are available in the sun and stars. For the proton-proton cycle to take place large reacting masses such as those we come across in the sun and stars are necessary. There may be considerable time interval between the initial fusion of two protons and their final incorporation in a α particle.

To adopt the fusion process, for production of energy on the surface of the earth, the reacting mass must be limited in size. Hydrogen bombs were made and tested on the earth. They are fusion bombs. The high temperature, needed to initiate the fusion reaction in a hydrogen bomb, is produced by the explosion of an atom bomb. There are many practical problems to be solved to achieve the production of power by fusion for peaceful purposes.

The fossil fuels (coal and mineral oils), which we use for power generation, will be exhausted after a few centuries. Fission presents a far greater energy source than the fossil fuels. Yet, there are certain disadvantages in the use of fission for power generation. Fission produces radioactive wastes the disposal of which is a serious problem. They can cause great harm to man if they are not properly disposed off. On the other hand, the products of fusion are almost harmless. Secondly, the fuel needed is available in plenty and may last for billions of years. In view of these advantages over fission, research on fusion is going on.

19.6 Radiation Hazards

After the invention of the atomic bomb, studies in the development of nuclear energy have gained utmost importance in many advanced countries. At the present time mankind is faced with the threat of atomic warfare. It is, therefore, important to know the extent of harm that penetrating high energy radiations can cause to human organism.

When a gamma ray or high energy nuclear particle passes through any material it ionizes the atoms it encounters on the way by knocking electrons from them. When an atom, in a complex organic molecule, is ionized the molecule may break up. The breaking up of some molecule could lead to the disruption of the normal functioning of the biological system. This might eventually result in the death of the organism. So, it is the degree of ionization produced, on passing through matter, that determines the strength of the radiation.

The dose of radiation that produces 1.6×10^{12} pairs of ions in 1g of air is accepted as the unit and it is known as the roentgen (r). The number of ions produced in living tissue by a dose of radiation of 1 roentgen is also of the same order of magnitude. One thousandth of the roentgen is the milliroentgen (mr).

The damage to the human organism, caused by penetrating radiation, depends upon (1) the dose, (2) the dose rate, and (3) the part of the body exposed. Hands and feet can receive a much larger dose of radiatoin without permanent injury than any other part of the human body.

The damage can be of two types :

(1) Pathological damage, and

(2) Genetical damage.

Pathological damage to the organism exposed to large doses of radiation could lead to the death of the organism. This will happen if a dose of radiation of about 600 roentgens is delivered over the entire body. Smaller doses of about 100 roentgens may produce leukemia and cancer which will ultimately lead to the death of the organism.

When nuclear radiations and X-rays pass through genetic cells they affect the chromosomes of the cellular nuclei. Injury to the genes in the reproductive cells is serious since it gives rise to mutations which are transmitted from generation to generation. Mutations are almost always harmful and the process is irreversible. It may be noted that, so far as inheritance of genetic damage is concerned, there is no safe lower limit of radiation. Only the probability of absorption of radiation by genes is small when the dose of radiation is small.

It is at present accepted that the safe limit for those exposed to radiation over the whole body is 250 mr/week. We are subject to radio-active radiations from the ground and by cosmic rays. The total amount of natural radiation received by a person from the time of conception till the age of forty is estimated to be about 4 roentgens. The additional radiation received during this period from medical X-rays, etc., may be another 4 roentgens. Due to this radiation. the resulting direct damage is negligible because the body has time to repair the injury.

At present the radiation due to radio-active fall-out caused by a atomic bomb tests is found to be only about 0.1 roentgen per person per year. This is small and if no more atomic bomb tests are going to be made. This dose rate will gradually decrease.

19.7 Accelerators

Soon after his discovery of artificial transmutation of elements, using fast moving α particles (See 18.8) Rutherford suggested to Cockroft and Walton to effect artifical transmutation in light nuclei using protons accelerated to very high voltages. A few years later they built a cascade generator for high voltages. Using these high voltages they accelerated protons and with them produced artificial transmutation in lithium nucleus.

Van de Graaff Generator

R van de Graaff extended the work of Cockroft and Walton by developing in 1931 the electrostatic accelerator which is called

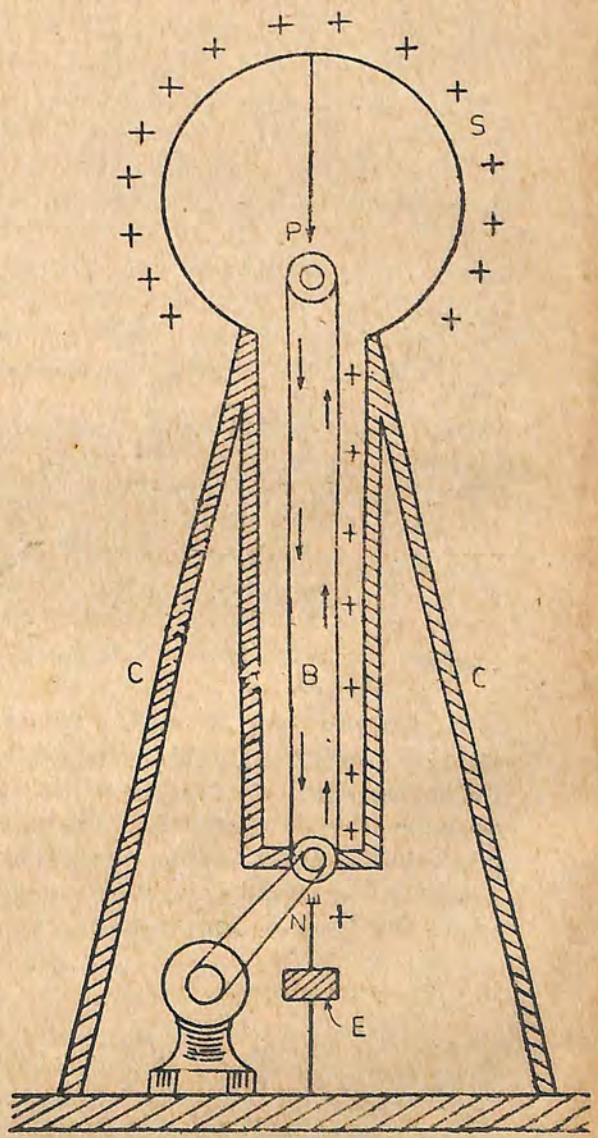


Fig. 19.6 Sketch of a Van de Graff generator.

by his name. This works on the principle that an electric charge imparted to a spherical conductor is distributed entirely on its outer surface.

The Van de Graaff generator consists of a large hollow sphere S supported on insulating columns C,C. (Fig. 19.6). A conveyor belt B conveys electrical charges from a battery at ground potential to the inside of the sphere. As the belt moves electrons jump from the belt to the row of needle points at N leaving the belt positively charged. After entering the sphere the electrons jump from the points P on to the belt leaving the outer surface of the sphere-positively charged. As more and more electrons are removed from the sphere positive potential rises higher and higher until a stage is reached when leakage into the surrounding air and through the insulator becomes significant.

By applying the high electric potential developed in this machine to one end of a long vacuum tube having an ion source, particles such as protons electrons or α particles are accelerated to energies of the order of a few million electron volts.

Cyclotron

Very soon it was realised that this method of artificially accelerating atomic particles to high energies has serious practial limitations. Soon afterwards a new concept of accelerating particles in circular motion, by application of smaller voltages, in repeated steps was conceived and this new type of accelerator known as cyclotron was designed and built by E.O. Lawrence at Berkeley, California University. In this device charged atomic particles are subjected to cyclic or circular motion in a chamber by applying a uniform magnetic field and the particles are accelerated once every half rotation by applying voltage of the order of a few hundred kilovolts across two cylindirical discs known as Dees.

The main part of the instrument consists of two short, hollow, half-cylinders D_1 and D_2 (Fig. 19.17). They are referred to as 'Dees' since they are in the shape of D.

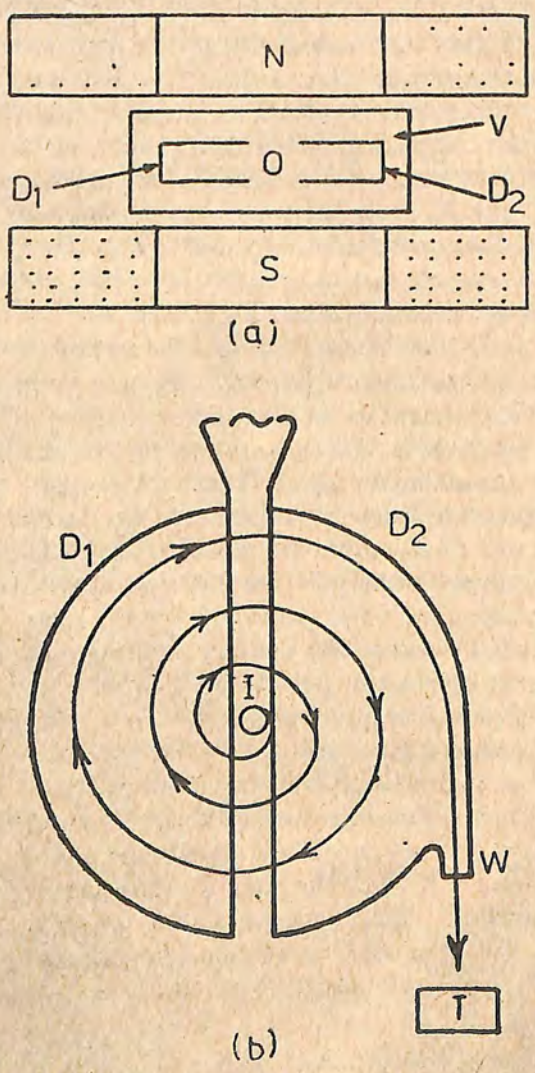


Fig. 19.7 (a) Section diagram of a cyclotron.
(b) 'dees' of a cyclotron showing path of a particle. Main parts of a cyclotron.

They are mounted inside a vacuum chamber V between the poles of a powerful electromagnet. The 'dees' are connected to the two terminals of a high frequency oscillator which supplies the energy to the dees. The entire device is in a high vacuum so that the molecules of air will not interfere with the acceleration of the particles.

Let us consider protons, for example, projected into D_1 by means of an ion gun I that works on the principle of the canal-ray tube. Since the particles move at right angles to the magnetic field they take up a circular path. Let the ions after moving through D_1 reach the space between D_1 and D_2. Suppose at this instant D_1 is positive and D_2 is negative. The electric field in the space between D_1 and D_2 accelerates the proton towards D_2. The accelerated proton moves faster through D_2 describing a larger semi-circle than before. As the particle just leaves D_2 the polarity of the Dees is automatically reversed by a properly tuned high frequency electronic oscillator. As such D_2 becomes positive and D_1 negative. The particle, therefore, receives a second push by the electric field in the space between the dees. It then moves through D_1 describing a larger semi-circle than before as it has now greater energy. Thus every time the proton reaches the gap between the dees the electric field is reversed automatically and the particle finally acquires high enough energy from the many small accelerated pushes it receives. The particle thus moves in a nearly spiral path as shown in (Fig. 19.7 (b)) and will be ejected through the window W by a deflecting voltage and hits the target T.

Theory : The particle moves in a circular path at right angles to the magnetic field. Since the magnetic field provides the necessary centripetal force, we have

$$Bev = \frac{mv^2}{R}$$

where B is the magnetic field, e is the charge on the particle, m its mass, v its velocity and R the radius of the circular path.
The time t for half a rotation is given by

$$\therefore v = \frac{e}{m} . BR$$

$$t = \frac{\pi R}{v} = \frac{m}{e} . \frac{\pi}{B}$$

t is, therefore, independent of both v and R. This fact is made use of in the above arrangement by making the time period of the high frequency alternating voltage equal to 2t.

By changing the frequency of the potential applied to the Dees and adjusting the magnetic field protons, deuterons and α particles of high energies can be produced.

Limitations : The expression for the time t taken by the particle to describe a semi-circle depends upon the value of m, the mass of the particle. We know, from Einstein's special theory of relativity, that mass m increases as the speed of the particle approaches the speed of light. If we try to produce higher speeds in a cyclotron, because of the change in the value of m, the value of t alters and this will throw the particles out of step with the oscillating field. To overcome this drawback many ingenious variations have been made in the construction of accelerators. We have as a result, the synchrotron and other machines. These machines employ changing oscillation frequencies or changing magnetic field strengths or both and can accelerate particles upto a few hundred Mev.

In India we have a variable energy cyclotron set up at Calcutta. This is designed to give 6 to 60 Mev Protons, 12 to 65 Mev deuterons and 25 to 130 Mev α particles.

This accelerator will be useful for studies in nuclear physics, solid state physics and radiation biology.

19.8 Cosmic Rays

Around the beginning of the present century, is was noticed that a charged gold leaf electroscope loses its charge even though it is mounted on an insulated support in a closed vessel. This was considered to be due to the conductivity of air caused by some unknown agency. Experiments proved that this was not due to radio-active materials of the earth. Even on lowering the electroscope into a lake it was found that the rate of discharge of the electroscope remained the same. Experiments using electroscopes carried in balloons upto 25 Km showed that the rate of discharge of the electroscope increased with altitude. In order to explain these observations Hess proposed a new hypothesis that the discharge of the electroscope is caused by a most penetrating radiation falling upon the earth from outside the earth. As the effect is shown both during the day and night, he concluded that this is not part of radiation from the sun but from outer cosmic space. These rays were, therefore, named as *Cosmic Rays.*

The gold leaf electroscope and other electrometers were used extensively in earlier studies of cosmic rays upto 1930. They were later replaced by the Wilson cloud chamber and the Geiger-Muller tube. In recent times photographic plates with emulsions specially sensitized for cosmic ray research have been used extensively, particularly for the study of cosmic rays at very high altitudes.

Primary Cosmic Rays and Secondary Cosmic Rays

Experiments have revealed that the cosmic rays entering our atmosphere are almost entirely composed of positively charged atomic nuclei. They are called *Primary Cosmic Rays.* About 89% of these primary cosmic rays are protons, 9% are rays and the rest are heavier nuclei like carbon, nitrogen, oxygen and iron. A high energy primary particle, after entering the atmosphere, soon collides with another atomic nucleus splitting one or both particles into a number of smaller nuclear fragments. These fragments share the high energy of the primary particle. They in turn, moving with high speed, collide with other nuclei and produce high-speed particles. In these collisions some elementary particles which are short-lived are also produced. During their decay some emit electrons and highly penetrating gamma rays. These photons and all the other high-speed particles that are produced, with the exception of the primary particle, are called *Secondary Cosmic Rays.*

At a height of about 25 km above the surface of the earth the number of secondary cosmic rays is about ten to fifteen times the number of primary cosmic rays that have entered the atmosphere As they move downwards many of the secondaries are stopped by collision because they lose energy by successive collisions. This energy is used up in heating the air molecules. A few high speed secondary cosmic rays reach the earth's surface. Some of them have enough energy left to penetrate through several hundred metres of earth and water.

Experiments on the latitude variation of cosmic ray intensity showed a minimum near the magnetic equator. The cosmic ray intensity at sea level is found to gradually increase on either side of the magnetic equator until a magnetic latitude of 42° is reached and thereafter it remains constant. This is due to the charged particles in the

cosmic rays being deflected by the earth's magnetic field. This does not mean that there is no cosmic radiation at the magnetic equator. Only a few high energy cosmic rays enter the magnetic equator.

Cosmic-ray showers : In 1928 Skobeltsyn, engaged in cosmic ray research photographed tracks in a Wilson cloud chamber. The tracks appeared in groups and all the particles in a group seemed to originate from some localized region near the wall of the cloud chamber. These are called *cosmic-ray showers*.

Fig. 19.8 shows the sketch of a cascade shower of cosmic rays photographed in a Wilson cloud chamber. There are some equally spaced lead plates across the cloud chamber. A high energy cosmic ray particles entered the solid block of matter and by collission with the nuclei of the atoms in the lead block secondary cosmic rays are produced. The tracks are the paths of these secondaries emerging out of the lower face of the metal block.

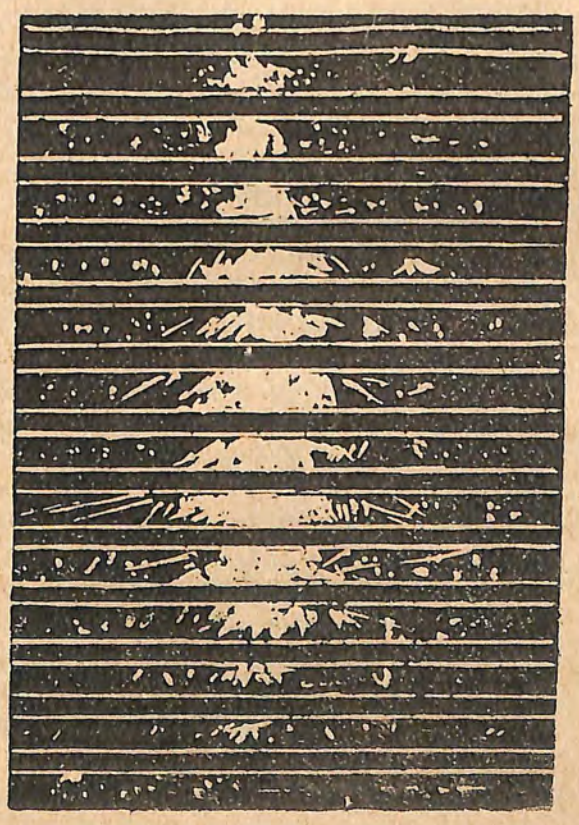


Fig. 19.8 Sketch of cascade shower of cosmic rays.

While studying the tracks of cosmic rays in a Wilson cloud chamber, Anderson discovered in 1932 a particle having the same mass as an electron but with an equal positive charge which is named as Positron. A positron does not exist for a long period in the free state. When a positron meets with an electron the two are annihilated. Positron is called an *Antiparticle* of electron.

Soon after the discovery of the positron, theoretical physicists began to work out the conditions under which a positron exists in nature. On the basis of the quantum theory of the electron they predicted that, if a high-energy photon (i.e. a high frequency gamma ray) were to come close to the nucleus of an atom, the gamma ray would be converted into a pair of particles, namely electron and positron. This is known as *pair production*. The experimental physicists in the field soon found such pairs of particles in a cloud chamber when Gamma rays were passed through matter.

Anderson and Nedermeyer in 1933

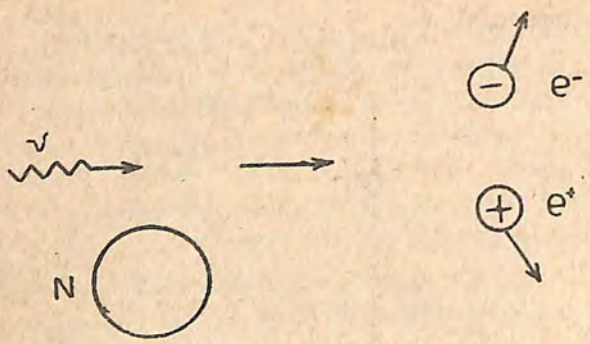


Fig. 19.9 Pair production—a high energy gamma ray comes close to the nucleus of an atom and is converted into an electron and a positron.

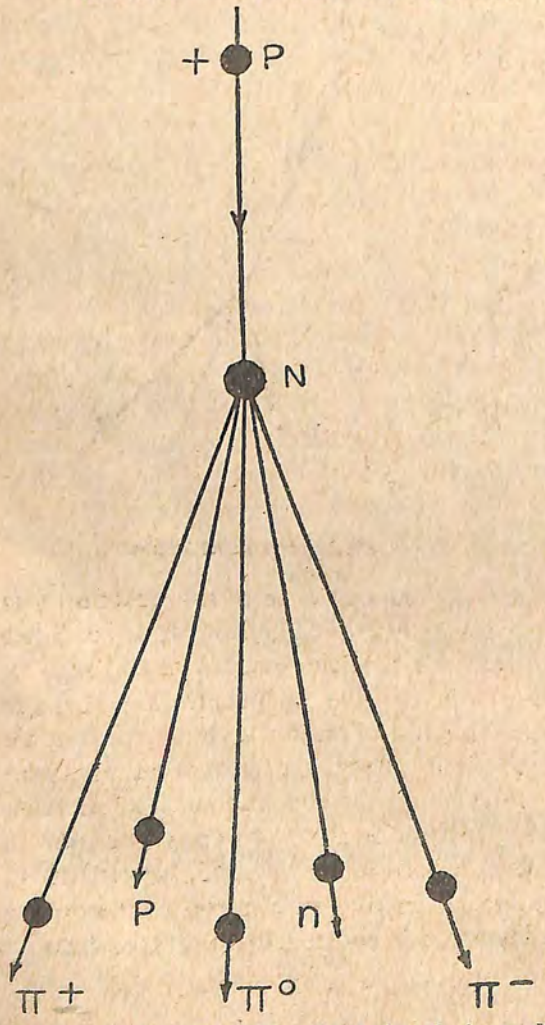


Fig. 19.10 A primary cosmic ray produced, by nuclear Collision, π mesons, protons and neutrons.

discovered the presence, in cosmic rays, of charged particles having mass about two hundred times that of an electron. They are produced by the collisions of cosmic rays with air nuclei. These particles are called mesons (See Section 19.9). They were first predicted theoretically by Yukawa in 1935.

Fig. 19.10 shows the production of π mesons or pions, protons and neutrons by a primary cosmic ray by nuclear collision. Some of the pions are positively charged, some negatively charged while some others are neutral.

All charged pions spontaneously disintegrate into a *μ meson* or muon and a *neutrino* (Fig. 19.11). The negative muon decays into an electron and two neutrinos. Neutral pions (π° mesons) are without charge and very unstable. They decay into two gamma rays. By electron pair-production these gamma rays create cascade showers of electrons in the upper regions of the atmosphere (Fig. 19.11).

19.9 Elementary Particles

From early times, natural philosphers believed that all material things are made up of a few basic particles. In their search for these fundamental or elementary particles, physicists have used atomic projectiles from powerful accelerators to bombard atomic nuclei. At the time of the discovery of the neutron, only four elementary particles were known. They are the electron, proton, photon and neutron. Of these particles, the electron, proton and neutron have definite rest mass but the photon has no rest mass. Subsequently more and more elementry particles were discovered. Some of them were predicted theoretically and confirmed later by experiment.

In radio-active decay it was found that the laws of conservation of momentum and

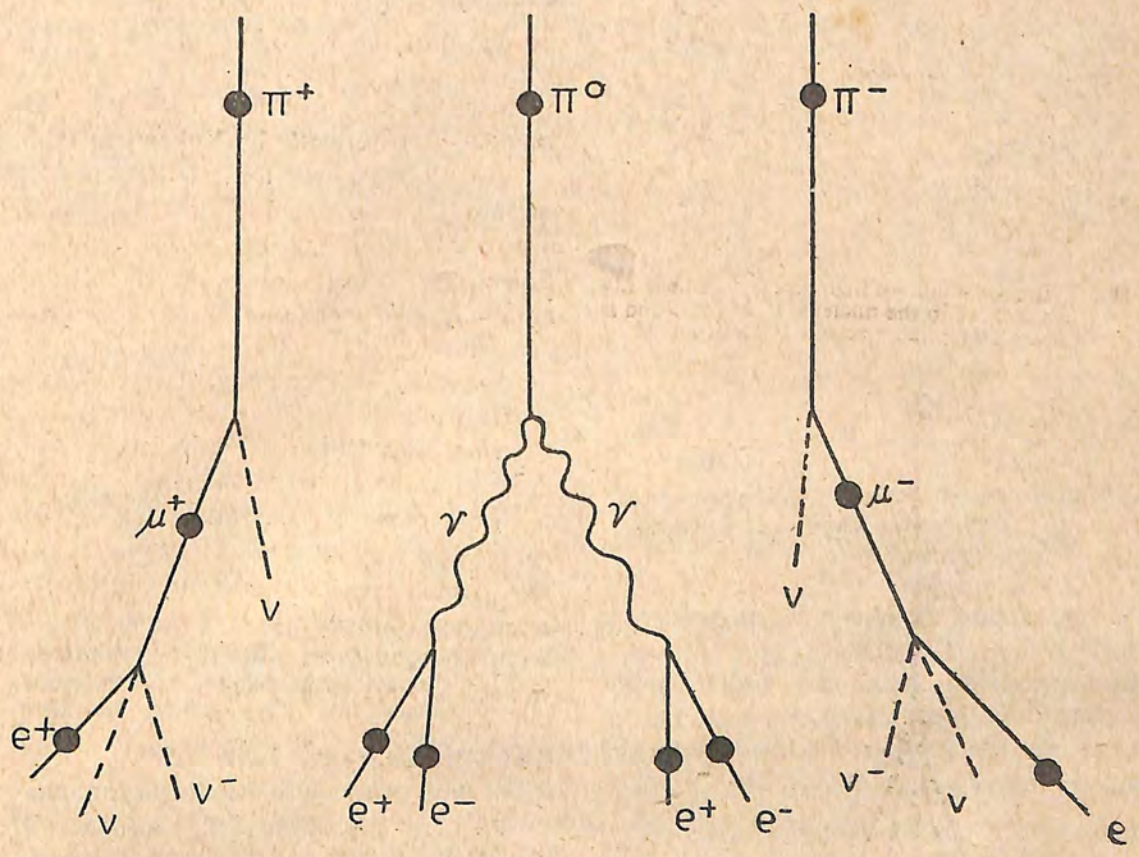


Fig. 19.11 Disintegration of π mesons into μ mesons, γ rays, electrons and neutrons.

energy were obeyed in the case of alpha and gamma decay. On the other hand, in beta decay a serious discrepancy was observed in that the laws of conservation of momentum and energy were not found to be obeyed. This discrepancy led pauli to postulate the existence of a new particle named neutrino which has no mass and charge but has angular momentum and energy. At the same time, it does not have the characteristics of an electromagnetic wave like photon. By postulating the emission of a neutrino along with the electron in beta decay process (Fig.19.12), it was observed that the laws of conservation of momentum and energy would be obeyed. The existence of the neutrino was experimentally proved in the year 1956. Since neutrino is a particle without any charge or mass it is capable of penetrating through large thickness of matter without appreciable interaction. They are generated in great abundance in the sun and in nuclear reactors. In beta decay in which a positron is emitted a neutrino is released, where as in beta decay involving electron emission an antineutrino is emitted, as indicated below :

$$p \rightarrow n + e^{+} + \quad + \upsilon$$

$$n \rightarrow p + e^{-} + \quad \upsilon^{-}$$

Earlier, in cosmic ray studies, reference

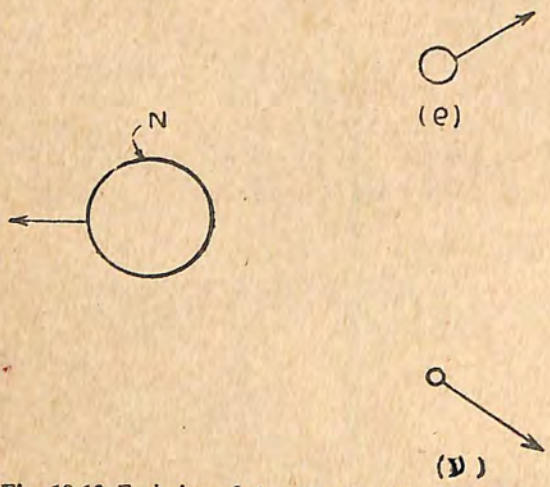


Fig. 19.12 Emission of electron (e) and antineutrino (υ) during beta decay, . . N—nucleus after β-decay.

was made to the discovery of the positron which is an antiparticle of the electron. When a positron comes into contact with an electron both get annihilated resulting in the emission of two gamma rays (Fig.19.13) each having an energy of 0.51 Mev.

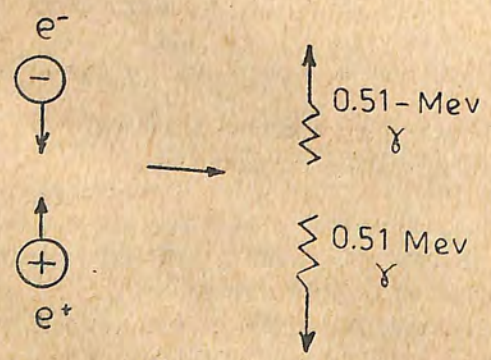


Fig. 19.13 Annihilation of positron and electron, when they come together, producing two gamma rays each having 0.51 Mev enery.

After the discovery of the positron the existence of a negative counterpart of proton was also predicted. Such a particle called the antiproton was experimentally discovered much later. When an antiproton comes into contact with a proton both are annihilated resulting in the production of a π meson with large energy. Sometimes instead of a π meson, two gamma ray photons, each having 936 Mev energy, are produced (Fig. 19.14).

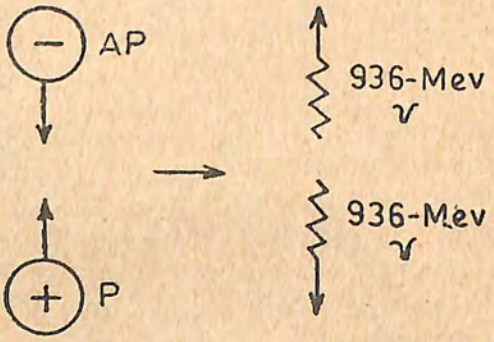


Fig. 19.14 Annihilation of a proton and anti-proton when they come together producing two gamma rays each having 936 Mev energy.

We have referred, in the section on cosmic rays, that two types of mesons, namely, π mesons or pions and μ mesons or muons were discovered in cosmic ray studies. A charged π meson has a mass of 273 m_e. Its life-time is 2.54×10^{-8} seconds. A charged π meson decays into a μ meson and a neutrino as indicated below :

$$\pi^+ \rightarrow \mu^+ + \upsilon$$

$$\pi^- \rightarrow \mu^- + \upsilon^-$$

There is also a neutral π meson which decays into two gamma rays

$$\pi^\circ \rightarrow \gamma + \gamma$$

The μ meson has a mass of 207 m_e and has a life-time of 2.22×10^{-6} seconds. The muons have great penetrating power because of their feeble interaction with matter. A

muon decays giving an electron and two neutrinos as indicated below :

$\mu^+ \rightarrow e^+ + \upsilon + \upsilon$

$\mu^- \rightarrow e^- + \upsilon + \bar{\upsilon}$

There is no neutral μ meson. Mesons are also classified as elementary particles.

Although, at one time, the total number of elementary particles was about thirty, recent researches have added more elementary particles to this list increasing the number to above 200. As the research work on elementary particles is advancing more and more, new particles with increasing complexity are being discovered.

Exercises

19.1 The isotope uranium $_{92}U^{238}$ decays successively to form $_{90}Th^{234}$, $_{91}Pa^{234}$ $_{92}U^{234}$, $_{90}Th^{230}$, and $_{88}Ra^{226}$. What are the radiations emitted in these five steps
($\alpha,\beta,\beta,\alpha,\alpha$)

19.2 If 200 Mev energy is released in the fission of a single nucleus of $_{92}U^{235}$, how many fissions must occur per second to produce a power of 1kw?
(3.12×10^{13})

19.3 Give an outline of the working of a simple type of nuclear reactor.

19.4 Assuming that in a star three alpha particles join in a single fusion reaction to form a $_6C^{12}$ nucleus calculate the energy released in this reaction, given mass of $_2He^4 = 4.002\ 604$ amu and that of $_6C^{12} = 12.000\ 000$ amu.
(7.3 Mev)

19.5 Assuming that four hydrogen atoms combine to form a helium atom and two positrons, each of mass 0.000 549 amu calculate the energy released.
(25.7 Mev)

19.6 Write a note on radiation hazards.

19.7 Describe the construction and explain the working of a van de Graaff generator.

19.8 Deuteron are to be accelerated in a cyclotron. The value of the uniform magnetic field applied is 2.475 Wb/m^2. What is frequency of the oscillating potential that must be applied to the dees of the cyclotron?
(mass of deuteron $= 3.3 \times 10^{-27}$ kg and charge 1.6×10^{-19}C) (19.0 MHz)

19.9 What is the period of revolution of a deuteron in a uniform magnetic field of 1 Wb/m^2. Use the data given in question no. 8. (1.296×10^{-7}s)

19.10 What are the primary Cosmic rays composed of ? What are the secondary Cosmic rays composed of ?

19.11 Why is the intensity of cosmic rays a minimum near the magnetic equator?

19.12 Draw diagrams to show the disintegration of Pions.

19.13 What is the antiparticle of the antineutron?

19.14 How much energy must a gamma ray photon have if it is to materialize into a proton-antiproton pair with each particle having a kinetic energy of 10 Mev? (mass of ${}_1p^1 = 1.007\ 276$ amu) (1896 Mev)

19.15 An electron positron pair is produced by a gamma ray of 2.26 Mev. How much kinetic energy is imparted to each of the charged particles? (0.62 Mev)

19.16 A 1 Mev positron collides head on with a 1 Mev electron and they are annihilated giving rise to gamma ray photons. What is the total energy of the resulting gamma rays?
(mass of electron 0.000 549 amu) (3.02 Mev)

UNIT 20

Electronic Devices

We live in an age of technology and one branch of technology that has revolutionised the present civilization is, of course, electronics. It is now possible for people to communicate speedily and efficiently with each other by using electronic appliances, such as radio, television, etc. which are the products of this technology. Similarly, radar and sonar and many other electronic instruments are used to defend our country from others. There are many other examples in our every day life that can be cited to show that electronics has now become an important tool of progress and welfare of the modern society.

Most interesting feature of all electronic devices is their speed of action. For example, the conventional method of communication between two individuals by letters etc. takes longer time in comparison with the almost instantaneous communication that can be achieved with the help of radio waves. Similarly, the time taken to multiply six figured numbers by conventional procedures may take many seconds while it takes only a few microseconds to do the same multiplication by an electronic calculator. It is because of such speeds, now achievable with electronics, that this technology has become so important.

20.1 Essentials of an Electronic Device

Electronics is essentially a branch of science which deals with use of electrons for specific applications. Electron as we know is the fundamental particle with the smallest mass and finite electrical charge. It is due to these two properties that electrons can be speedily and easily manipulated. Manipulation of electrons can be done either in the form of their number or their positions. We have already studied the effect of magnetic and electric fields on electrons and are quite aware that by applying such fields it is possible to change the direction of their motion conveniently. Moreover, the energy that is required to manipulate this way is very small. At the same time, the velocities attained by the electron in the process are very large and hence account for high speed of action. All electronic devices essentially make use of these important properties of electrons.

It is now clear from the above discussion that electronic devices must therefore have the following three parts :

1. a source of electrons
2. an arrangement for manipulating them, and
3. an arrangement for observing the

desired effect due to the manipulation.

In this unit we will see that these three features are essential to all electronic devices.

20.2 Source of Electrons

Since electrons form a major constituent of atoms which constitute the matter they are available in large numbers. They however, tend to interact with matter that comes in their contact due to their peculiar electrical nature ; and form bonds. This tendency, in fact, accounts for the difficulties in freeing them from substances, particularly, the insulating materials which are the most abundant form of matter around us. It is, therefore, necessary to look around for these materials in which the binding of electrons is weak enough to allow for easy removal for further use from their bonding atoms.

(a) *Metals as source of electrons*

In the chapter on solids we can see that metals contain freely moving electrons which account for the high electrical conductivities of the metals. These electrons are, however, confined to the conductor and cannot leave its surface at the nominal temperatures or electric fields. It is easy to understand if we consider what happens at the moment, the electron decides to leave the surface. The removal of the electrons immediately produces positive charge on the surface of the conductor which was electrically neutral before the electron came, just out of the surface. The positive charge therefore, pulls back the electrons that try to esacpe. Thus the electrons in the metal which happen to come on the surface do not get opportunity to leave and hence, remain virtually confined to it. This situation is very similar to that of molecules of water trying to escape against forces at the surface in the process of evaporation. In the case of electrons, however, the pull at the surface is governed by the nature of the material. The energy that is required to remove the electrons from the surface so as to overcome the retaining forces at the surface, is called the work function and is measured in electron volts (eV). Typical values of work function for various materials are given in Table 9.2. It is seen that the energy needed to remove electrons from tungsten and molybdenum is higher than that required for their removal from barium and cesium.

It is now obvious that if we choose to use electrons which are easily available in metals, some methods of overcoming the pull at the surface must be devised. One method which can be easily thought of is, of course, to subject the metal to a very high electric potential which will extract the electrons from within and make them free. This type of removal of electrons by applying high electric fields is called field emission. Since the use of high voltages is technically difficult and often dangerous, the field emission is not used as a method of obtaining electrons in appliances often used in the home.

While studying the physics of atoms, we have already studied photoelectric effect and the laws of photoelectric emission. Photoemission from materials can also act as a source of electrons and has been conveniently used in some special devices. In this case, however, the energy of photons required to release the electrons from the metal surface must be greater than the work function. Both the photo emission and the field emission have limited applications

because the rate of removal of electrons, i.e., the current available from these sources is generally very small.

There is another method of getting electrons in large numbers and this has been in use for a number of years. In this method a metal filament is heated by passing enough current through it. Increase in the temperature of the filament increases the thermal energy of the electrons which in turn helps them to overcome the pull at the surface. At sufficiently high temperatures, a reasonably high number of electrons find themselves with energies larger than the work function and leave the filament. This type of emission is called the *thermionic emission*.

The electrons emitted from the surface of the filament are not directly available for manipulation but tend to remain more or less in the space near the filament in the absence of an electric field, and form a cloud known as *space charge*. Since like charges repel each other, this cloud of electrons tends to force back to the filament other electrons that are being emitted. The amount of electron emission is thus governed by the space charge that the cathode has around it.

It is obvious that, in order to obtain substantial emission of electrons, the metal should be heated to as high a temperature as possible. Therefore, it is customary to use metals having high melting points such as tungsten or molybdenum as materials for heaters. Being metals themselves, these heaters can provide emission of electrons. However, their work functions being higher, they are not efficient enough for large emission currents. In order to improve the efficiency of emission, tungsten heaters are generally coated by materials like barium oxide which have low work functions. Such high efficiency heaters are used in vacuum tubes.

(*b*) *Semiconductors as source of electrons*

So far we have seen that to obtain electrons for devices, we need to remove them from the metals by various techniques, into the free space for manipulation. These methods have technical disadvantages such as operation at high temperature, provision of energy for heating etc. Moreover, since the electrons are free, they have to be manipulated for the purpose of the device only in an evacuated space, where they would not collide with gas particles which would only make manipulation difficult. The device therefore becomes bulky.

Since the last three decades we have been using devices made of semiconductors which do not have all these unfavourable characteristics. These devices are small and do not require additional heating, and hence can work at room temperatures. This is possible because semiconducting materials can provide free electrons which can be manipulated in the material itself without bringing them into the free space. Thus these devices are very small, hardly a square mm in size. We have already studied in the Unit on Solids, the way electrons behave in semiconductors. Here we will consider only those aspects which are relevant to electronic devices.

Germanium and silicon are the semiconducting elements most commonly used in devices. In the crystalline form, their atoms are placed in the diamond like structure, and bonded covalently with each other. The bonding electrons i.e., the outermost electrons of these atoms, are bound with so small an energy that it is possible even at normal room temperatures to free

some of them from their atoms. Once freed they can move about easily until caught by another atom which had already lost its electron. The number freed, however, will depend on their bond energy in the semiconductor. For example, silicon which required higher energy for dislodging an electron from its bond gives less number of free electrons for a given temperature compared to germanium which has correspondingly lower bond energy. This accounts for the fact that majority of semiconducting devices are made of silicon which is least affected by temperature.

Let us now see what happens if we apply electric voltage across such a semiconducting crystal. Fig 20.1 shows a piece of germanium connected to a battery, a switch and an ammeter to record current that would flow in the circuit. In fact, when

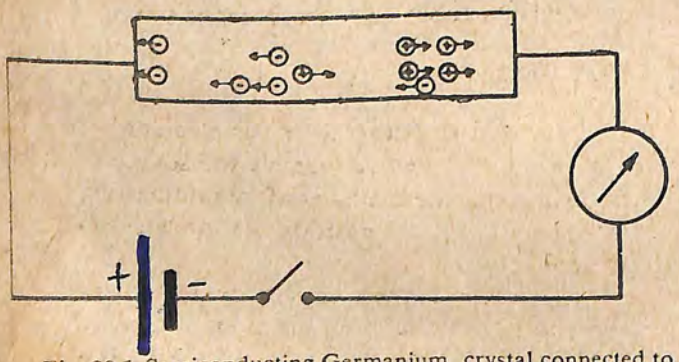


Fig. 20.1 Semiconducting Germanium crystal connected to a battery.

the switch is put 'on' the meter would show current flowing through the circuit. This current most obviously be due to charges moving in the semiconductor. This is hardly surprising, considering the fact, that the experiment is done at room temperature i.e., about 300K which provides enough thermal energy to release electrons and move under the influence of the applied electric field and register a current. We have already learnt in the unit that this current is composed of two opposite currents produced by 'n' type carriers, i.e., electrons and 'p' type carriers i.e., holes. They are produced in equal number to account for the electrical neutrality of the material. This type of material, which has to be very pure, is called intrinsic semiconductor. If we increase the temperature of the semiconducting piece, say, by holding it over a flame, current in the meter is seen increasing very rapidly. When the semiconductor is cooled to the room temperature, the current drops to its original value. Those charge carriers can also be produced by exposing the semiconductors to light, however, unlike photoemission in metals, in which it escapes the metal surface, the electron released by the striking light remains within the semiconductor and hence can be easily manipulated. For an efficient device, it is necessary to have a source of electrons which can give ample but controllable number of electrons. Intrinsic semiconductors, in that respect prove to be a poor source of electrons, and technically difficult to control or manipulate. Instead, by adding impurities of the right type, called dopants, at the time of preparing crystals of semiconductors it is possible to enhance the number of either 'n' or 'p' type charge carriers in them. Semiconductor that gives 'n' type carriers is called an 'N' type semiconductor while that which gives 'p' type charge carriers is called a 'P' type semiconductor. It is easy to see that by controlling the amount of dopant, we can control the strength of our sources of charge carriers and fabricate them to suit the requirement of the device to be made.

It is not always necessary in practice, to use only doped crystals. Since the size of the actual crystal is small, the dopants can be made to enter the desired regions of a

pure crystal by depositing them on the surface and diffuse the dopant atoms at higher temperatures. This makes fabrication of semiconducting devices rather simple.

20.3 Manipulation of Electrons

So far we have studied various sources of electrons or charge carriers which can be used for making electronic devices. As has been mentioned earlier, these electrons are manipulated in the device to produce the desired effect. It is, therefore, obvious that the choice of method of manipulating electrons will depend mainly upon the final effect we are looking for. In this section therefore we will try to recapitulate some of the methods of manipulating electrons which have been studied earlier in the course.

There are two ways of manipulating electrons. One simple way is obviously to change their numbers, while in the other the number of electrons is kept fixed but their direction of motion in the device is changed to produce the desired effect. The first method involving the number of electrons is easy to understand because we know that in most of the electrical devices, currents can be increased by increasing the voltage applied to the device and vice-versa. For example, electrons when flow through a resistance will give rise to a current as a result of the voltage applied to it and according to the Ohm's law if the voltage is doubled the current is also doubled and so on. Therefore, control on the number of electrons can be achieved by merely subjecting the device to appropriate voltages. Such devices are called linear devices because they behave according to a linear relation such as Ohm's law.

The second method of manipulation, however, depends on changing the original path of the electron to make it reach a different position so that the desired effect can be obtained. We have already learnt that if a beam of moving electrons is subjected to an electric field, acting perpendicular to the path of the electrons, then it deflects in the direction of the electric field. The resultant displacement is decided by the force acting on the electrons in the direction of the field and its original velocity. Thus by changing the field in a desired manner, it is possible to manipulate the position of an electron. The other method of deflecting electron beam in space is by subjecting it to a magnetic field again perpendicular to the original direction of motion. The resultant shift in the path will then be governed by Lenz's law. These two methods of manipulating the positions of electrons are effectively used in devices such as television tubes and oscilloscope tubes.

20.4 Diode

Now that we know the basic elements of any electronic device, namely the source of electrons and the methods of manipulation, we are in a better position to understand their functions.

The simplest type of electronic device constitutes an electron emitting source on oxide coated tungsten filament, called a cathode, and a surrounding metal electrode to collect electrons called anode or plate, both enclosed in an evacuated envelope. Such a device is called a diode (di means two in Greek and ode is from the word electrode). The structure of a diode is shown in Fig. 20.2. We have already learnt that the diodes are used as elements to rectify a.c. voltage i.e., to convert a.c. voltages into d.c. voltages. The process of rectification demands that the diode allows

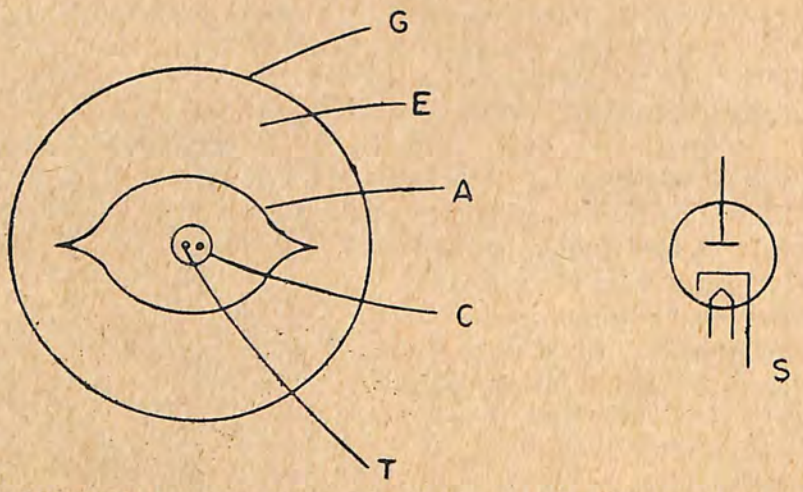


Fig. 20.2 Diode G—glass envelope, E—evacuated space, A—anode, C—cathode, T—tungsten filament, S—symbol for diode.

current to be drawn only in one direction on application of an alternating voltage. To be precise, we can say that when the anode is positive relative to the cathode, the electrons from the cathode should flow in one way only from the cathode to the positive anode. On reversing the polarity of the voltage, there would be no flow of electrons. In other words the diode can be considered as a sort of switch which is 'on' when the anode is positive relative to the cathode and opens when the respective polarities are reversed. Since electrons are used for such a switching it is an extremely fast acting switch.

(*a*) *Working of a diode*

If a positive voltage is applied to the anode, the electrons from the space charge cloud which are close to it are first attracted giving rise to the current. However, the removal of the electrons from the space charge will allow equal number of electrons to come out of the cathode surface and maintain the number of electrons in the space charge the same. The electrons which were attracted to the charged plate would neutralise the positive charge as shown in the circuit in Fig. 20.3. A current is registered due to the passage of the charge back to the cathode. Internal flow of electrons from the cathode to the anode is called *space current*. In this type of device the anode current is equal to the space current.

The space current as it is easy to see, will flow only in one direction from the anode to the cathode. Since the plate is not a source or emitter of electrons, but merely collects them, there is no charge available to flow from the anode to the cathode in the event of the reversal of the polarity of the voltage.

(*b*) *Diode characteristic*

Let us now see how the anode current behaves when the anode voltage is changed. As the positive voltage on the anode is gradually increased more and more electrons

from the space charge are taken up by the anode and therefore the current increases with the voltage. After a certain voltage the space charge is completely removed and the electrons, emitted by the cathode, travel directly towards the anode so that the rate of release of electrons by the cathode is equal to that at which the plate collects them. Increasing anode voltage beyond this point will, however, not result in an increase in the current, since entire emission of the cathode has been taken up by the

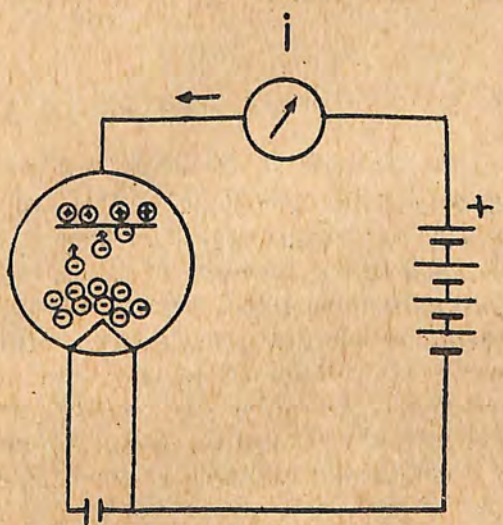


Fig. 20.3 Diode circuit, i—space current.

anode. This limit of current that can be reached is called saturation current. The dependence of anode current on its voltage is shown in Fig. 20.4.

(c) Diode as a rectifier

Suppose we now replace a battery in Fig. 20.3 by a generator of a.c. voltage The common a.c. voltage available from the domestic supply consists of a voltage varying sinusoidally at the rate of 50 times a second. Fig. 20.5(a) shows such a voltage applied to the anode. The a.c. petential here swings in one direction for half the cycle and in

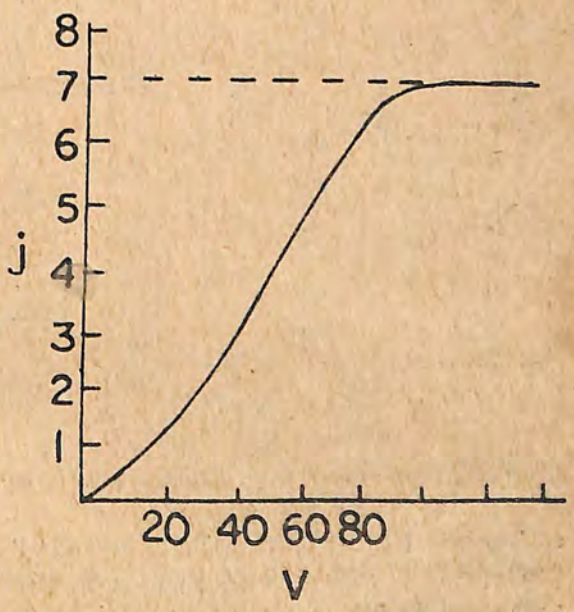


Fig. 20.4 Diode characteristic.
V—anode voltage, j—current.

opposite direction for the remaining half. If we assume that on application of such a voltage, the first half of the cycle alone makes the anode go positive with respect to the cathode, then the current in diode will flow from the anode to the cathode. However, on reversal of polarity due to the later half of the cycle, the diode will not pass any current. The nature of the current is shown in Fig. 20 5.(b), which essentially is a pulsating current, but remains on one side of the zero axis, all the time. It, therefore, can be considered as a direct current (d.c.) with added variation in its amplitude. This process of changing alternating current to direct current is called rectification and in majority of cases, the diodes are used only for this purpose. The rectifying action of the diode in relation to its characteristics can be visualised by using a graph shown in Fig. 20.6(a). Here we have really combined two graphs in one since they are related to

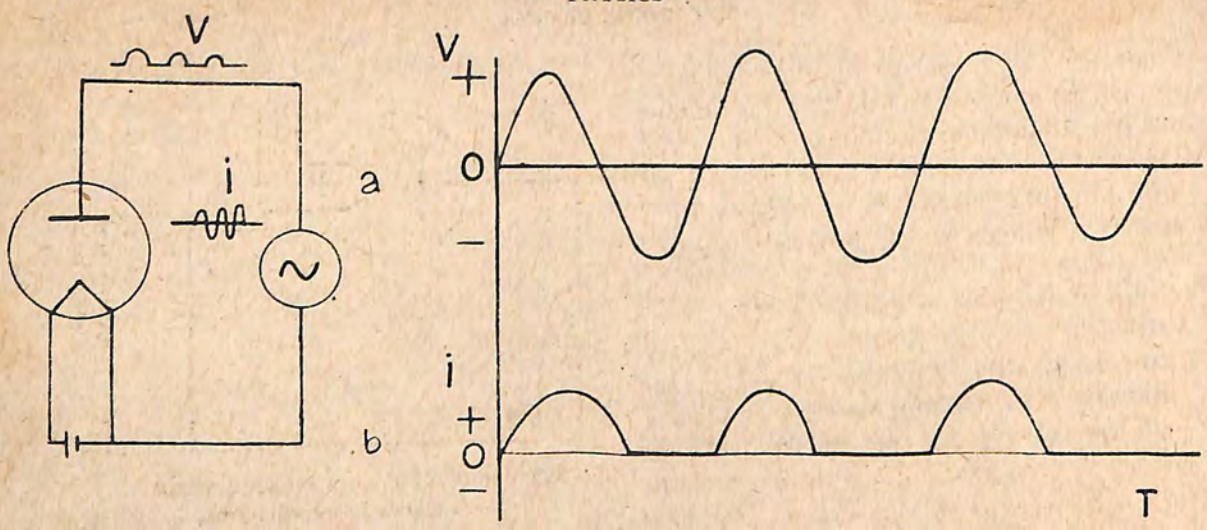


Fig. 20.5 Rectification of A.C. Voltage applied to a diode. V—anode voltage, i—anode current, T—time. (b) Rectified current.

each other. The vertical axis here is extended downward to represent time and is used to show sinusoidal variation of the anode voltage. The sinusoidal variation is shown here to have been broken into four equal intervals. On the horizontal axis, we again have time to show the resulting current governed by the diode characteristics indicated by the pulsating d.c. output current which appears only in the positive half cycles. There is an advantage in showing the behaviour of the diode in this fashion. It will be obvious that if we apply higher and higher a.c. voltage to the diode at some voltage the diode will reach a point where it cannot deliver more current because of the limits imposed by the saturation current, as shown in Fig. 20.6(b). It is, therefore, obvious that for proper operation of a rectifier it is necessary to stay within the

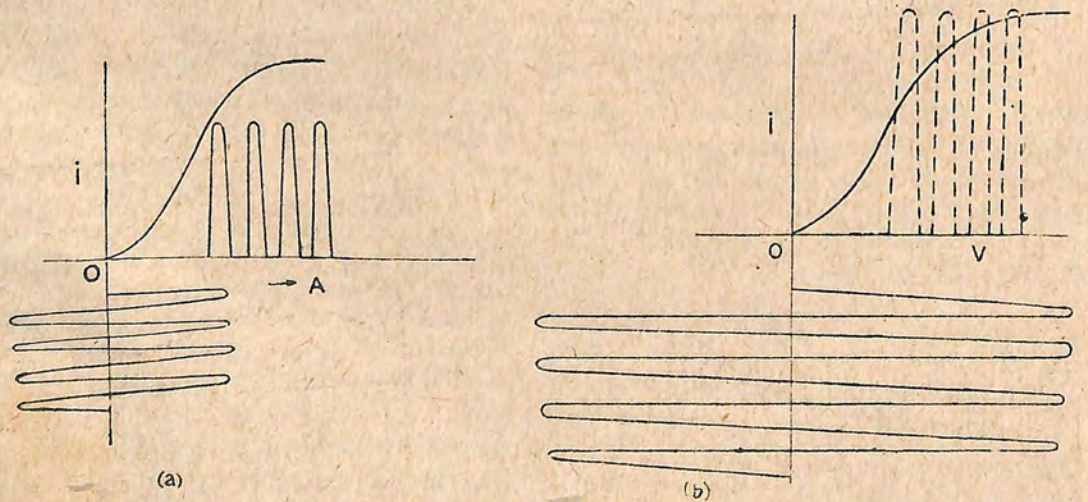


Fig. 20.6 (a) Rectification action of a d ode in relation to its characteristics. V—anode voltage i—anode current. (b) Saturation.

limits of voltage specified by the design of the diode. It is also clear that the voltage current relationship of a diode is no more linear and hence such a device will behave as a non-linear device in contrast with a simple resistance which follows the linear Ohm's law.

In practice such a pulsating d.c. is of very little use. It is, therefore, necessary that before putting it to use some arrangement must be made to remove the pulsations and extract the d.c. part of this voltage. This is done by smoothing circuits consisting of capacitors and inductors. Fig. 20.7 shows a circuit in which the capacitor acts as a reservoir of charge, delivered to it by the diode, and helps to nearly remove the pulsations in the voltage across the resistance connected across it. This is a simple method of developing d.c. voltages which can be used as a source of power for electronic circuits.

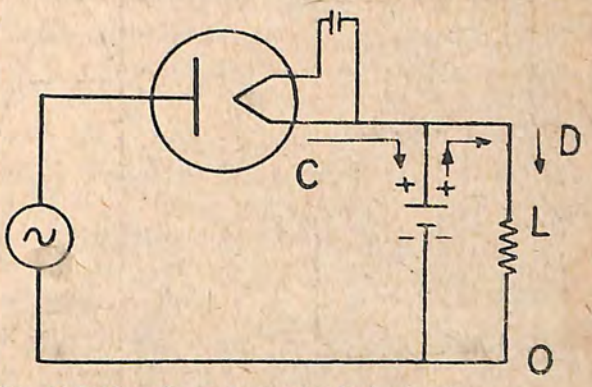


Fig. 20.7 Use of a capacitor in diode circuit
C—charging, D—discharging, L—inductor

(*d*) *Diode as a detector*

We have already learnt that for long range communication with radio waves it is necessary to modify a sinusoidal radio frequency in such a way that it can carry audio signal on it. This process of impressing a signal upon a carrier is called modula-

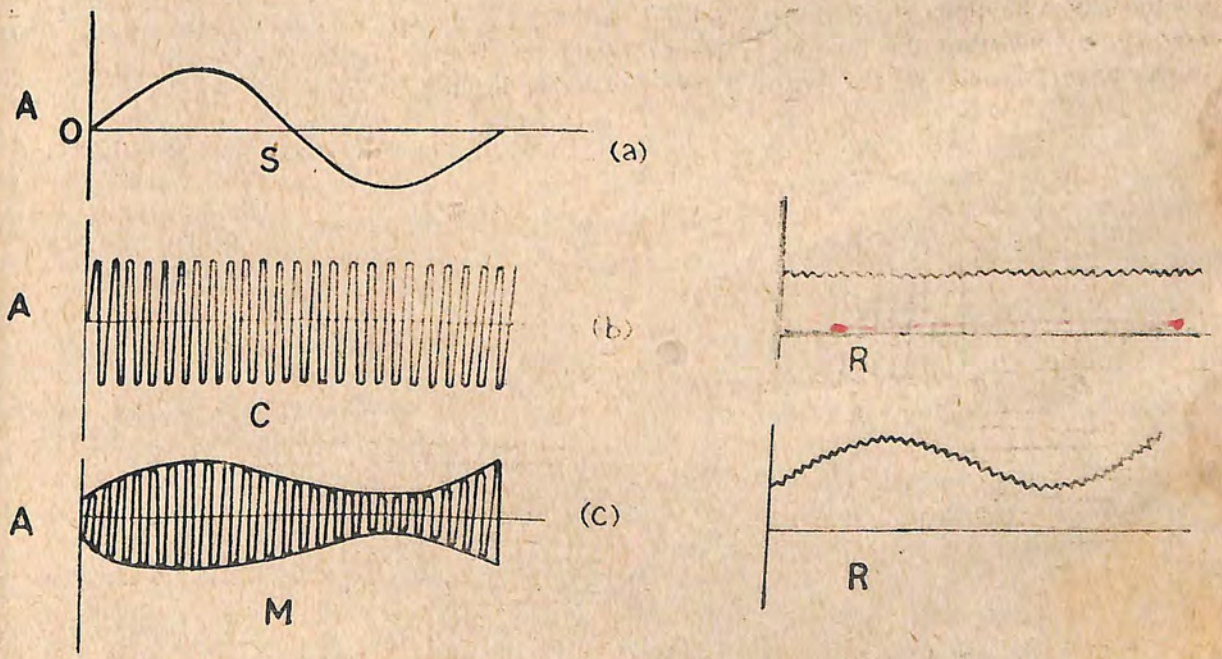


Fig. 20.8 Modulation of a sound signal, S—audio signal, C—carrier wave, M—modulated wave, R—rectification, A—amplitude.

tion. Let us consider a relatively simple case of a transmitter, transmitting a signal which consists of a single frequency Fig. 20.8(a). The transmitter will generate a suitable carrier wave as shown in Fig. 20.8(b). In the transmitter the sound signal and the carrier are mixed to produce a modulated radio frequency carrier, by a method in which amplitude of the carrier is made to vary in direct relation to the audio frequency. Such a wave is called amplitude modulated wave and is shown in Fig. 20.8(c). The original audio frequency wave is now on both the sides of the carrier.

Let us now apply the carrier voltage in Fig. 20.8 (b) to a diode rectifier circuit as had been done earlier in Fig. 20.7. Since the amplitudes of all the waves of the carrier are equal, on rectification and proper smoothing, we would obtain a d.c voltage proportional to the amplitude of the carrier wave. Suppose that we now apply the voltage of Fig. 20.8 (c) to the same circuit, since the amplitude of each carrier wave cycle is varying in accordance with the impressed audio signal, we will get on rectification, a voltage which will be similar to the audio signal. This is shown in Fig. 20.8. Thus the diode can be used to demodulate the modulated wave or detect the signal from the carrier.

20.5 Semiconductor Diodes

So far we considered the devices which used electrons moving from one electrode to another on application of electric field, by making one electrode a source of electric charge and the other a mere collector. We have seen that the current in these devices can flow only in one direction. In the past two decades semiconductor diodes are largely used in electronic circuits. In a normal piece of semiconductor it is not possible to make one part of it behave as a source of charges and another as a collector. One has to artificially generate conditions such that one part of the piece acts as a source of electrons while the other would be deficient in electrons and therefore would collect them on application of an electric field.

(*a*) *p-n Junction*

In is only under this condition that such a system will behave in a way similar to a diode. In practice this is achieved by forming a junction between an 'N' type crystal to a 'P' type crystal as shown in Fig. 20.9. It is uneconomical to use two separate crystals to form a junction, instead by a suitable technique, regions having the necessary N or P type of material are created side by side in the same piece of semiconductor, thus forming a p-n junction.

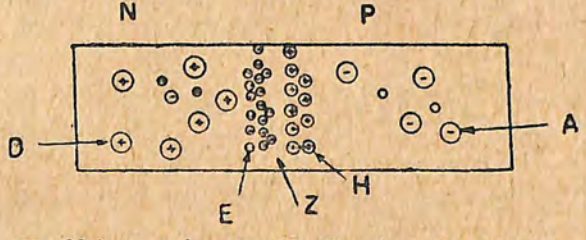


Fig. 20.9 p—n junction N—N region , P—P region, D—ionised donor, A—ionised acceptor, H—hole, Z—depletion zone, E—electron.

Let us now see the behaviour of charge carriers in a p-n junction. In the 'N' region of the crystal we have greater concentration of electrons than that in the 'P' region. Hence some of the excess electrons of the impurity atoms, called the donor atoms, in the 'N' region will drift in to the 'P' region after breaking away from their parent atoms. After losing the electrons the donor atoms in the 'N' region, will be-

come positive ions. Since they are fixed in place within the crystal framework at the lattice points, the positive ions thus formed remain fixed in the 'N' region near the junction. The 'N' region near the junction has, therefore, lost some of its electrons and is no longer neutral; it is positively charged. Similarly holes from impurity atoms, called acceptor atoms, in the 'P' region would drift towards the junction giving rise to the negative ions. As a result the 'P' region close to the junction similarly loses electrical neutrality and is negatively charged.

The electrons which have diffused into the junction from the 'N' region meet holes from the 'P' region at the junction. Being oppositely charged, they recombine thus removing both the carriers. However, as more holes from the 'P' region are forced to diffuse across the junction they are repelled by positive ions in the 'N' region. Similarly the electrons from the 'N' region which attempt to diffuse across the junction towards the 'P' region are repelled by negative ions of the 'P' region This fact gives rise to accumulation of -ve charges at the junction on the side of the 'N' region and similar accumulation of +ve charges on the side of the 'P' region with a very thin separation between them. Such an accumulation of opposite charges on two sides of the junction results in creation of an electric field across it, called a barrier potential. It is called so because the charges must overcome a barrier like potential to cross the junction. Its strength is decided by the nature of the original material from which the junction is prepared. For example, in the case of silicon it is of the order of 0.6 volts and somewhat less in the case of germanium. This barrier stops further diffusion of the electrons from the 'N' region and of the holes from the 'P' region. The zone within which this barrier potential is developed is called the 'depletion' zone.

(*b*) *Junction rectifier*

Let us now consider what happens if a p-n junction is subjected to an external electrical field supplied by a battery, as shown in Fig. 20.10. The negative terminal of the battery is connected to the 'N' region and the positive terminal to the 'P' region. This arrangement of applying voltage to a junction is called forward bias.

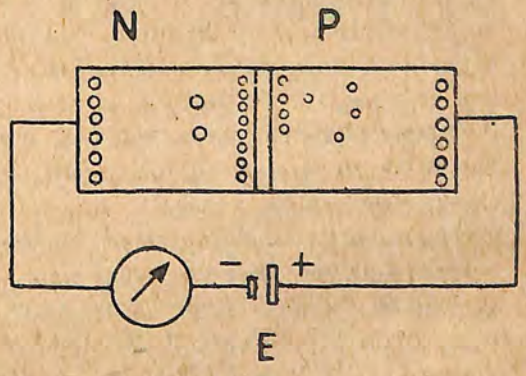


Fig. 20.10 p-n junction connected to a battery (forward bias) N—N region, P—P region.

On application of this electric field, the electrons of the 'N' region are repelled from the negative terminal of the battery and drift towards the junction. Similarly, the holes from the 'P' region would drift towards the junction because of the positive charge at the other end. This will force carriers to move towards each other thus reducing the width of the depletion zone. In addition the external field being in opposition to the barrier potential, the strength of the barrier potential is reduced.

Once the strength of the barrier potential has been reduced the electrons are now able

to cross the junction meeting the holes which have crossed the barrier to enter the depletion zone in a similar way and recombine with them. These recombinations of electrons and holes is registered as current in the external circuit. As long as the external battery is applied this flow of charges will continue to take place thus giving rise to the forward current. Although the height of the potential barrier originally was a few tenth of a volt, it is often seen that voltages as high as one or one and half volts are required to overcome the barrier. The voltage excess over the original potential barrier is due to the drop that takes place in the body of the crystal, forming the depletion region. Having once overcome the barriers, charge carriers will flow giving rise to the forward current which will increase with increasing voltage. However, it will be accompanied with an increasing energy loss in the body of the crystal due to the voltage dropped in the other parts of the crystal. Such an energy loss appears as increase in the temperature of the device which is detrimental to the functioning of the diodes. The relationship between the forward currents and the applied voltage for a p-n junction diode is shown in Fig. 20.11. Note here that the current does not start flowing for voltage smaller than that required for overcoming the potential barrier.

Let us connect the external battery in the opposite manner such that the positive terminal is now connected to the 'N' region and the negative terminal to the 'P' region. In this case it is said to have a reverse voltage or bias applied to the p-n junction. A careful observation of the way the charge carriers would now behave shows that the electrons from 'N' region will be attracted towards the positive end of the battery,

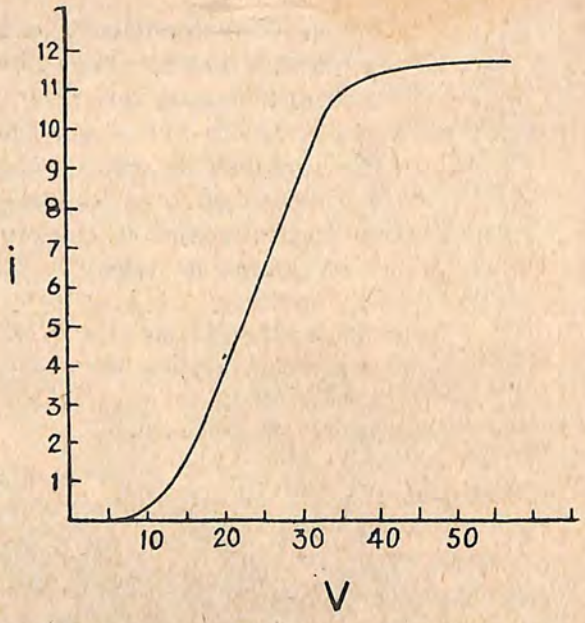


Fig. 20.11 Relationship between forward current and applied voltage for a p-n junction.

and the holes will similarly be attracted, towards the negative end of the battery, thus keeping the carriers away from the depletion area. In this case obivously no current will flow, since on application of electric field in this manner, the barrier potential will be more than what it originally was. Therefore, in principle no current should flow in a reverse bias p-n junction. However, at room temperature, there is always a possibility that some of the donors and acceptors which are near the depletion zone give away charges giving rise to a recombination and hence a very small reverse current. In fact we have seen that p-n junction can act in a way similar to the diode.

Such junctions are now extensively used in transistor radios, televisions and many other applications of electronics.

20.6 Triodes

In the diodes so far studied by us, the

manipulation of electrons to produce affects like that of rectification was done merely by controlling the applied external electric fields to its two electrodes. It is possible to have a better control on the electrons by providing an additional electrode for manipulating their numbers. There are two types of devices using this technique of manipulation. The first type is known as a triode (*tri* for three and *ode* from electrodes) in which electrons are controlled in vacuum and second is the transistor, which similarly controls charge carriers in semiconductors.

(*a*) *Construction of a vacuum triode*

First we will look into the working of a vacuum triode and see how the manipulation has been done in a more refined fashion for totally different purpose compared to that in the diode. Let us insert another electrode in between the cathode and the anode of the diode, after taking care to see that the electrons can pass through the new electrode. It is possible to do so by providing a wire mesh which will permit us to establish the necessary electric field for manipulation but will allow the manipulated electrons to pass through its structure. This electrode, is, therefore, called grid. In actual triode this grid consists of spiral or an open mesh of circular or elliptic cross section and is made of a thin iron, molybdenum or special alloy wire. For efficient control and collection of electrons, the filament or cathode is held at the centre and the grid is kept close to the cathode surrounding it. Both of them are then surrounded by a cylindrical anode for efficient collection of electrons. The structure of the triode is shown in Fig. 20.12.

(*b*) *Working of triode*

In the absence of any voltage applied to the grid with respect to the cathode, i.e. when the grid and the cathode are connected together, the triode obviously behaves like a diode. However, if the grid is given a negative

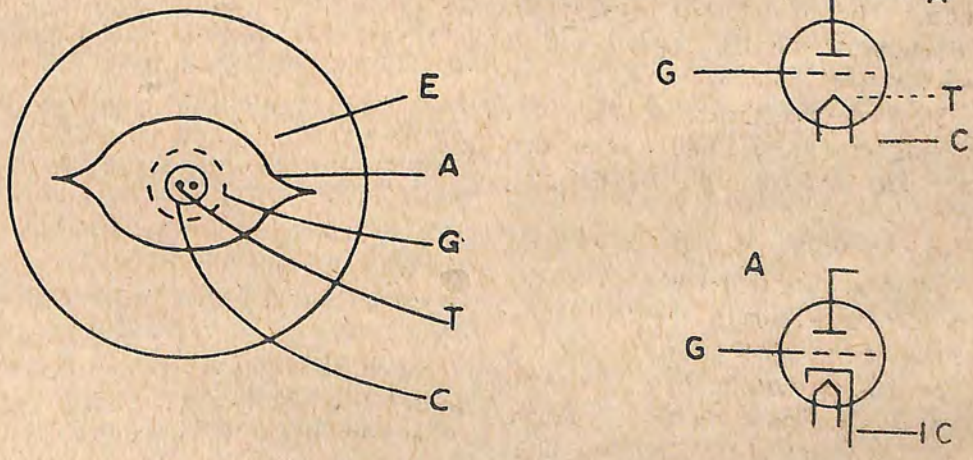


Fig. 20.12 Details of triode. Directly and indirectly heated cathode, E—evacuated space, A—anode, G—perforated Grid, T—tungsten filament, C—cathode, I C—indirectly heated cathode.

charge with respect to the cathode it will establish an electric field which will force the electrons from the space charge of the cathode to move towards the cathode. In that case, the number of electrons that can pass across the grid to the anode which is already held positive will obviously go down. For example, for small negative charge on the grid, many electrons will still reach the anode, and if it is increased, fewer electrons can get through, thus reducing the current. If the grid is made negative so as not to allow any electron to pass through, the space current will obviously be zero. On the contrary, if the grid is given a positive charge, electrons will be pushed rapidly towards the anode, giving rise to an increase in the current. If the grid is made positive enough it will only assist to remove electrons from the space charge to the anode and the triode will reach to the point where the currents cannot be increased further. That is the rate at which the anode would remove the electrons is equal to or greater than the rate at which the electrons are emitted by the cathode. Hence, at higher voltage anode takes all electrons from space charge so that the electrons emitted from the cathode can go straight to the anode. This is similar to the sturation point which arises in the case of a diode.

To use the triode it is necessary to provide, firstly, currents for its heater so that the cathode can function effectively as a source of electrons. Secondly, it will also need a voltage to make the plate positive with respect to the cathode. This is known as a plate voltage supply and thirdly in order to establish the desired electric field between the grid and the cathode an additional supply called grid cathode supply is also required.

In actual practice, it is convenient to consider the cathode. the source of electrons as a reference point with respect to which all potentials could be applied. A potential of a few hundred volts is often used between the anode and the cathode. The grid is, however, kept at a potential of a few volts of negative polarity with respect to the cathode. The actual value of the grid voltage is chosen in such a way that the triode draws approximately half the current for the given anode potential, at zero grid potential. A circuit of a triode using these voltages is shown in Fig. 20.13.

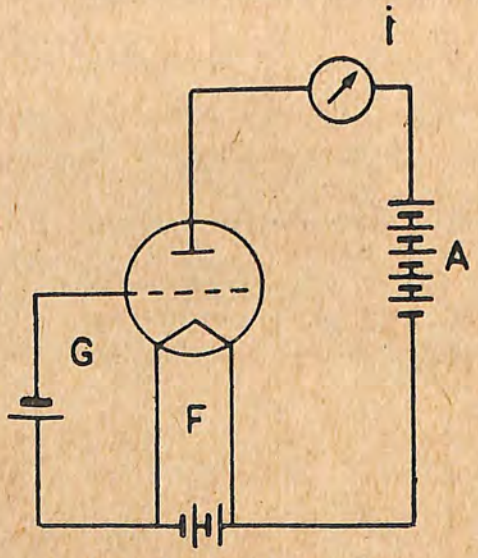


Fig. 20.13 Triode circuit.

(c) Triode characteristics

In the case of diodes, because we have just two electrods, we had one characteristic curve which corresponded to the dependence of anode current on the anode voltage. In the case of the triode, due to the presence of the third electrode i.e, the grid, we also need to take into consideration the dependence of the anode current on the voltage impressed on the grid. The anode current depends on the voltage impressed between the grid and the cathode, and also on the

voltage between the anode and the cathode. These three parameters namely, grid voltage, anode voltage, and anode current are thus interrelated. Their dependence on each other is measured by keeping one constant while the other two are changed, and, in fact, we need minimum two relations to find all the three dependences. For example, we examine variations of the anode current as a function of the anode voltage keeping the grid voltage constant. We can also study the relationship between the anode current and the grid voltage keeping the anode voltage constant. This relationship, also known as triode characteristic, forms the basis for the use of the triode in an actual circuit.

Let us first consider the triode circuit in which we have held the grid voltage at a fixed value. If we now change the anode voltage over a range say, from zero to 500 volts, we will get a change in anode current as the voltage changes. This is given in Fig. 20. 14. The figure also contains similar currents for different grid voltages thus forming a family of curves, called anode characteristic curves. Though the triode is a three terminal device, by holding its grid at a fixed voltage, we have in effect made it behave like a two terminal device. Each anode characteristic curve thus gives the relationship between the applied voltage V_a and resultant current I_a through this two terminal device. We can recall that two terminal electrical element with a similar behaviour is, of course, the electrical resistance. The anode characteristic therefore helps to find the effective resistance, called *anode resistance* r_p of the triode. This is the internal resistance of the device for the given

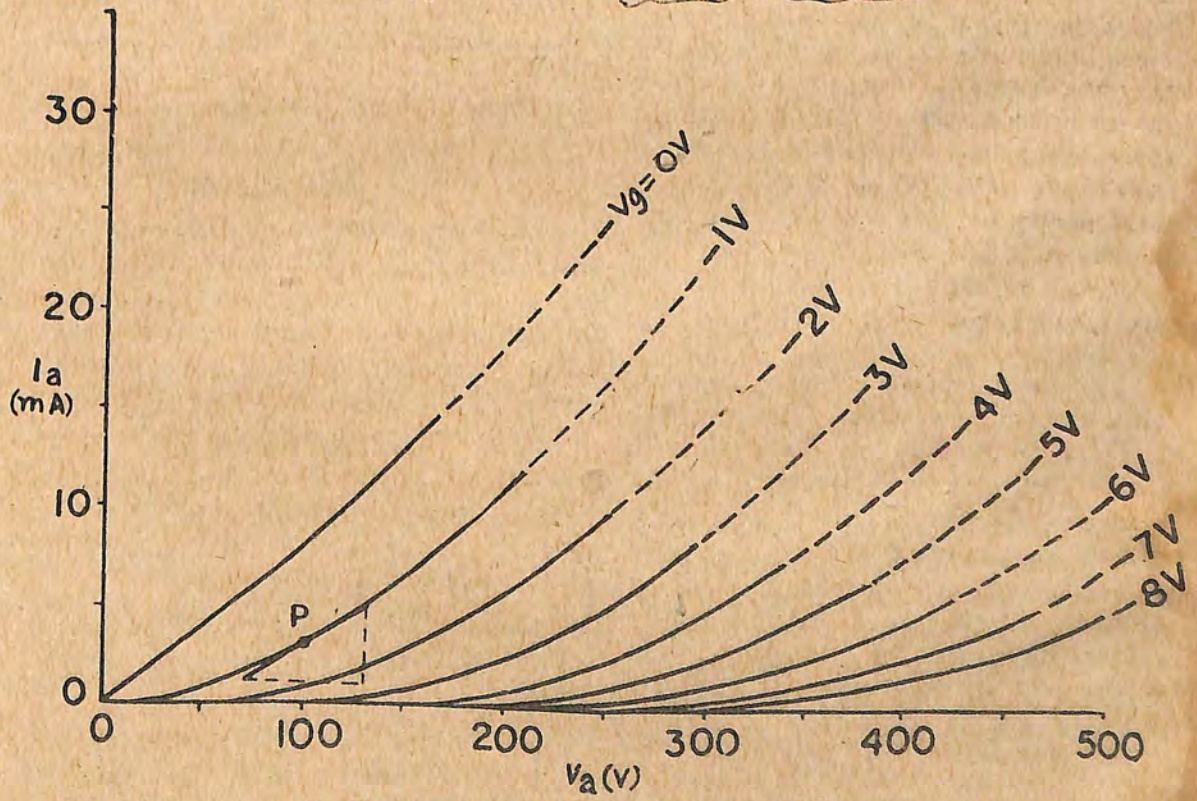


Fig. 20.14 Anode voltage and anode current characteristics.

grid and anode voltage and can be obtained from the slope of the plate characteristic curve. The anode resistance r_p is given by

$$r_p = \left(\frac{\delta V_a}{\delta I_a}\right)_{V_g=\text{const.}} \text{ ohm} \quad ...(20.1)$$

In a similar manner, we can examine the relationship between the changes in the grid voltage with the resultant changes in the anode current, when the anode voltage is kept constant. The circuit to be used for such a study is the same as the one given in Fig. 20.13. Let us start by fixing the anode voltage at say 100 volts and apply to the grid negative voltage, say, −3 volts. The meter will register the anode current corresponding to these voltages. As the grid voltage is made less negative, more anode current will be registered. As the gird voltage becomes more and more negative the anode current will continue to fall until it becomes zero, after which any increase in the negative grid voltage will not help to change the anode current. The grid voltage at which the anode current reaches zero is called the 'cut off' grid voltage of the triode. We can repeat this process and obtain the set of curves shown as in Fig 20.15, for different values of the anode voltage It can be observed from the curves that higher the anode voltage used, greater is the cut off grid voltage. This family of curves is called mutual characteristic curves.

In these curves, we are really dealing with the same two terminal device, but the current I_a flowing in the device is related to the voltage V_g on the third electrode. The slope of these curves gives a measure of the manipulation of the current flowing through the device by the grid voltage, called the transconductance g_m. It is given by the following relation.

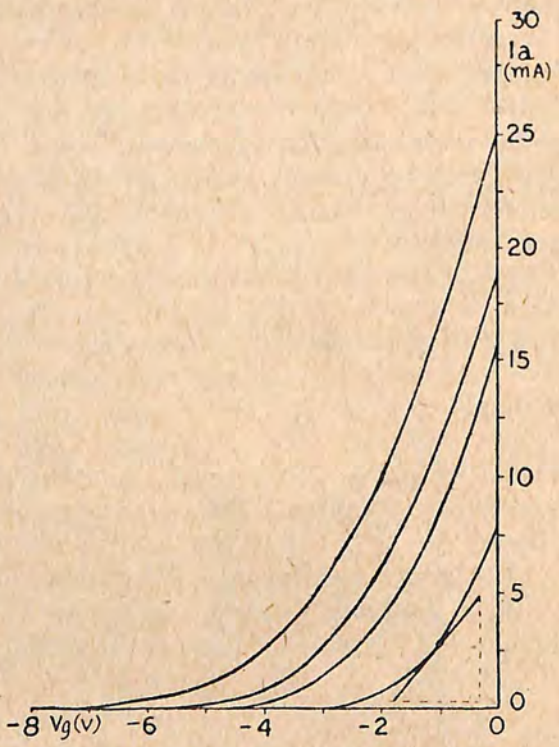


Fig. 20.15 Mutual characteristic curves of a triode.

$$g_m = \left(\frac{\delta I_a}{\delta V_g}\right)_{Va=\text{constant}} \quad S^*...(20.2)$$

The third characteristic describing the dependence of the anode potential on the grid potential for constant anode currents is not very much used, since in actual practice, the two parameters, r_p and g_m are adequate. However, this dependence can be easily derived from the following relation :

$$\left(\frac{\delta V_a}{\delta V_g}\right)_{I_a} = \left(\frac{\delta V_a}{\delta I_a}\right)_{V_g} \times \left(\frac{\delta I_a}{\delta V_g}\right)_{Va}$$

Here, $\left(\frac{\delta V_a}{\delta V_g}\right)_{I_a = \text{constant}}$ is the variation in anode potential due to that in the grid potential and is called the amplification factor μ.

We thus have a useful relation :

$$\boxed{\mu = r_p \times g_m} \quad ...(20.3)$$

* S (Siemens) is the SI Unit of Conductance formerly known as mho.

The triode can therefore be considered as a source of potential which generates potential equal to μ times that applied to the grid. The potential of this source appears across the anode and the cathode. As every source of potential, this one has its internal resistance equal to r_p, whose value is decided by the d.c. potential applied to the electrodes of the triode. We will later see how this aspect is used in understanding the applications of the triode.

EXAMPLE 20.1

Find the values of r_p, g_m and μ for the characteristics of the triode (ECC81) shown in figure 20.14. The operating conditions of the triode are :—

$$V_a = 100V ; I_a = 3 \text{ m A} ; V_g = -1V.$$

Solution

By the equation 20.1

(a) $$r_p = \left(\frac{\delta V_a}{\delta I_a}\right) V_g = \text{constant}$$

This is the slope of the anode characteristic of the triode at the given V_g.

Now for the given conditions ; i.e. at $V_g = -1V$.

By drawing the tangent at point P in the figure.

For $I_a = 3$ mA on the curve for $V_g = -1V$

$\delta V_a = (130 - 60) = 70V$;

$\delta I_a = 5 - 1 = 4$ m A.

$$r_p = \frac{70}{4 \times 10^{-3}} = 17.5 \text{ K}\Omega$$

(b) To find g_m; by Equation 20.2 $g_m = \left(\frac{\delta I_a}{\delta V_g}\right) V_a = \text{constant}.$

Consider the mutual characteristic i.e., $I_a - V_g$ curves.

For $V_a = 100V$ and $I_a = 3$ mA, which corresponds to the point P of the above example.

$\delta I_a = 5 - 0 = 5$ mA.

$\delta V_g = 1.9 - 0.3 = 1.6$ V.

$$g_m = \frac{5 \times 10^{-3}}{1.6} = 3.2 \text{ mS}^*.$$

(c) To find μ ; consider Fig. 20.14. We have

$$\mu = \left(\frac{\delta V_a}{\delta V_g}\right) I_a = \text{constant}$$

For $I_a = 3$ mA, we draw a line parallel to V_a axis to intersect the curves corresponding to $V_g = 0$ and $V_g = -2$. The intercepts give

$\delta V_a = 152 - 40 = 112$ V

$\delta V_g = 2 - 0 = 2V.$

$$\mu = \frac{112}{2} = 56$$

Now $\mu = g_m \times r_p$ at the given point

$g_m = 3.2$ mS $r_p = 17.5$ K

$\mu = 3{,}2 \times 10^{-3} \times 17.5 \times 10^3$

$\mu = 56$

Thus the relation holds good.

20.7 Transistor

Earlier we saw that by inserting an electrode, a wire mesh, between the plate and the cathode, we could control the flow of electrons, thus creating a very useful device called the triode. One wonders whether a simple modification of a p-n junction diode can be made so that the current across the junction can be controlled as efficiently as was done in the case of the triode. Such a device will neither need an evacuated envelope nor heated source of electrons, which makes the device bulky.

Fig 20.16 shows a device called transistor, which behaves very much like a triode. It consists of a 'P' type thin material with 'N'

* S (Siemens) is the SI Unit of Conductance formerly known as mho

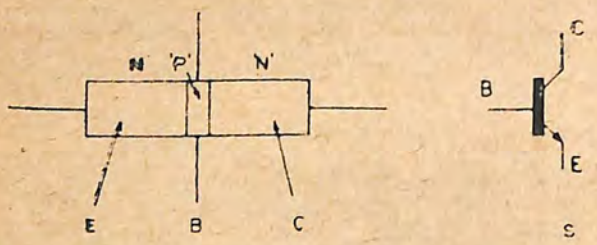


Fig 20.16 Transistor E—emitter, B—base, C—collector, S—symbol.

type materials on its either sides. These are bonded to form n-p-n system with two p-n type junctions connected in opposition. This is then equivalent to two p-n diodes connected back to back. However, since we want a source of electrons for the device, one of the 'N' region will have excessive donor concentration compared to the other. This region is known as *emitter*, and the other 'N' region as *collector*. The 'P' region forming junctions on both of its sides with them is called *base*, which, in a manner similar to the grid in a triode, controls the passage of electrons from the emitter to the collector. In fabricating the device, care is taken to make the base thickness as small as possible, about a few microns, so that it only controls the flow of electrons without causing their loss, while crossing the base region to reach the collecter.

If we keep in mind the functioning of p-n junctions when subjected to external voltages, it is easy to see that one of the two potential barriers will increase in its strength and widen, and hence will prohibit flow of electrons from the emitter to the collector. Thus, in the absence of any voltage applied to the base, i.e., base is kept unconnected, the transistor will not conduct at all. Let us, on the other hand, consider an arrangement shown in Fig. 20.17 in which a positive voltage is applied to the base with respect to

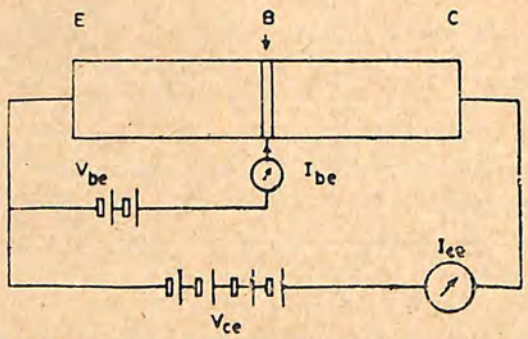


Fig. 20.17 n-p-n Transistor Circuit E—emitter, B—base, C—collector.

the emitter. As a result, the strength of the potential barrier at the junction between the emitter and the base is reduced and the electrons are let into the base region. Their number will, of course, depend on the voltage in such a way that higher this voltage larger the number. At the same time the strength of the barrier at the base collector is also lowered, so that the electrons can pass across to the collector. The base is so thin that the electrons do not stay too long to recombine with the holes. In fact, very few electrons, about 1 to 5 per cent recombine giving rise to the current I_{be} in the base-emitter circuit. This current can now become the measure of the ability of the base to manipulate the current I_{ce} in the collector emitter circuit. Thus we now have a device to which electrons from a source are let to a collector in a controlled fashion, an action similar to that of a triode. The parameter which quantitatively expresses efficacy of the control is given by

$$\beta = \frac{I_{ce}}{I_{be}} \qquad \ldots(20.4)$$

This relation is known as 'd.c.' current gain of a transistor. Transistors are designed with β values ranging from 20 to 200.

If the base is biased in reverse direction i.e., the base is made more negative with respect

to the emitter, then the strength of potential barrier at the base emitter junction is enhanced thus stopping the electrons from moving towards the collector. The transistor in that case is 'off' and is equivalent to an open switch in the circuit. This is similar to a triode biased to 'cut off'. The variation of I_{ce} with the collector voltage Vc shown in Fig. 20.18, for various values of I_{be}.

We have so far discussed a n-p-n configuration of a transistor. It is easy to see that transistors with a p-n-p configuration will behave in an identical fashion, except that the polarities of voltages applied to the electrodes will have to be opposite to that used for an n-p-n transistor. The transistor

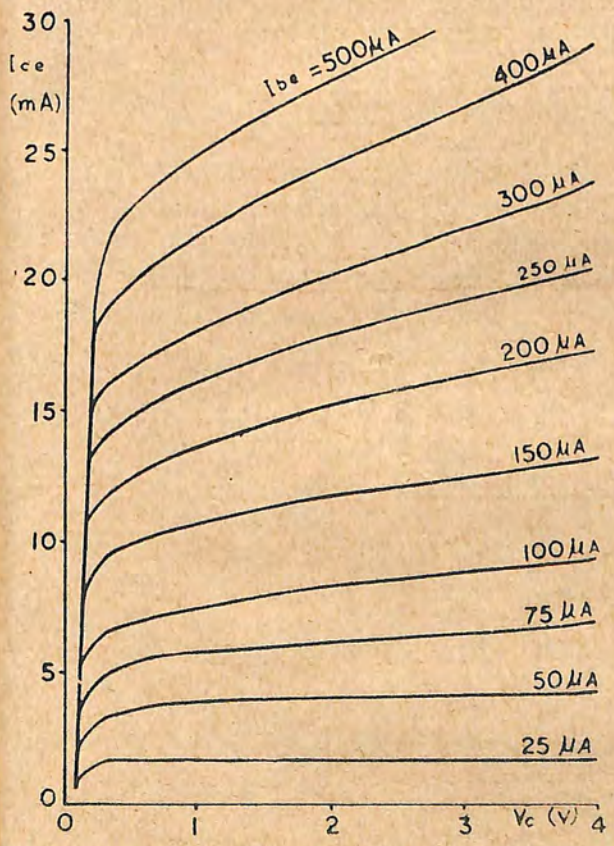


Fig. 20.18 n-p-n transistor charecteristics.

can thus be considered as a source of current, delivering current I_{ce} which is β times the current flowing in the base emitter circuit. If the current in the base-emitter circuit is changed by δI_{be} the generator current will change by δI_{ce} which is equal to $\beta\delta I_{be}$. The generator will, of course, have its internal resistancer due to the resistance of the semiconducting material of the transistor. We will use this aspect later in understanding the application of a transistor.

EXAMPLE 20.2

Find the value of the d. c. current gain, β for the transistor (BF 115) from the transistor characteristic given in Fig. 20.18, for $V_{ce} = 2V$.

Solution:

By equation 20.4 $\beta = \dfrac{I_{ce}}{I_b}$

For $V_{ce} = 2V$, from the given characteristic.

$I_{be} = 25\ \mu A$

$I_{ce} = 2.5$ mA

$$\therefore \beta = \frac{2.5 \times 10^{-3}}{25 \times 10^{-6}} = 100$$

At any other point say at $I_b = 50$ Micro ampere

$I_{be} = 50$ Micro ampere

$I_{ce} = 4.9$ mA

$$\beta = \frac{4.9 \times 10^{-3}}{50 \times 10^{-6}} = 98$$

At $I_{be} = 200\ \mu A$

$I_{ce} = 15.8$ mA

$$\beta = \frac{15.8 \times 10^{-3}}{200 \times 10^{-6}} = 79$$

20.8 Applications of Triodes and Transistors

So far we have shown how the first two characteristic, i.e. the source of electrons and an arrangement for their control has been made in simple devices. We will now see how the third and last feature namely, the method to obtain the desired effect after manipulation.

(a) *Amplifiers*

Often we find microphones, record player heads and many other electrical devices produce such feeble electrical voltages that it is necessary to magnify or amplify them by a large factor before their use to operater, say, loudspeakers. Amplifiers are electronic circuits which accept small voltages and amplify them say 'A' times. 'A' is called 'amplification factor or gain' of the amplifier. Triodes and transistors, which give large variation in the currents flowing through them for small control voltages, can be suitably used as devices for amplification. We will, however, have to make some arrangement externally to make use of these currents in such a way that the resultant or output voltage is many times greater than the control or signal voltage.

Fig. 20.19a shows a practical circuit for a triode amplifier in which the varying anode current is made to pass through a resistance R_L by connecting it between the anode and the supply voltage. In this circuit, the signal potential e_g to be amplified is applied to the grid. We have already seen that a triode can be considered as a source of potential generating $\mu \times e_g$ volts and that it has an internal series resistance of r_p ohms. The circuit in Fig. 20.19a can be replaced by an equivalent circuit shown in Fig. 20.19b. It can be seen from it that

$$E_o = \frac{\mu\, e_g \times R_L}{R_L + r_p} \qquad ...(20.5)$$

$$\text{Amplification factor } A = \frac{E_o}{e_g} = \frac{R_L \mu}{R_L + r_p}$$

The output voltage will depend on the value of R_L only for the given operating d.c. potentials. For $R_L = \infty$, the gain of the amplifier is equal to the gain of the device.

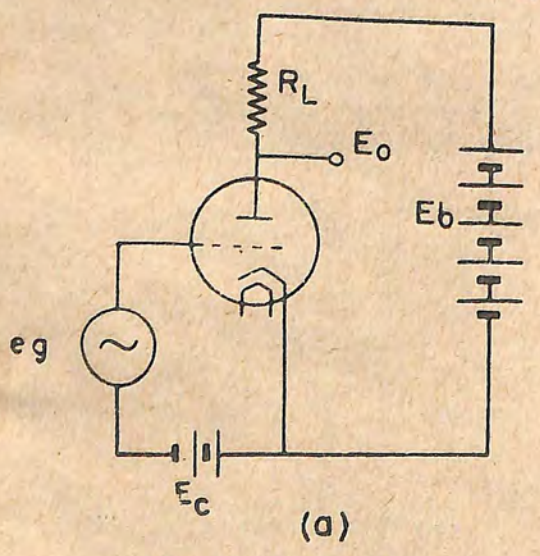
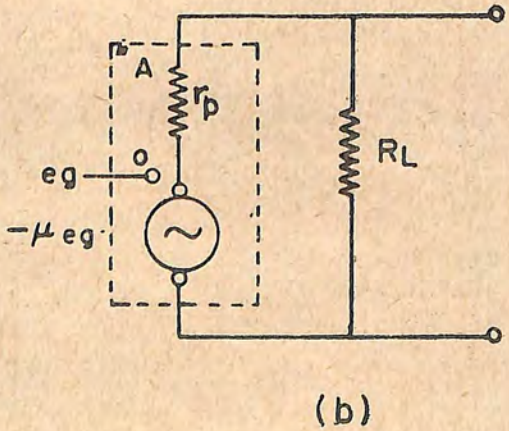


Fig. 20.19 (a) Triode amplifier.



(b) Equivalent Circuit.

Thus the practical amplifiers will always have gain less than device gain.

For example for $R_L = r_p$, $A = \frac{1}{2}\mu$. The practical values of μ generally range from 20 to 100 and hence the circuit will function as an amplifier.

Fig. 20.20 similarly shows a practical circuit for a transistor amplifier. The varying current I_{ce} is, in this case, generated due to the passage of current I_{be} in the base emitter circuit as a result of an application of a signal to the base.

Amplifiers form an important basic units of most of the electronic circuits in use. It is therefore customary to indicate amplifiers by a symbol given in Fig. 20.21 which shows all its essentials such as its gain, input, output and supply terminals.

It is obvious from the circuits that the amplifiers in fact use the energy provided by the battery supply E_b for their

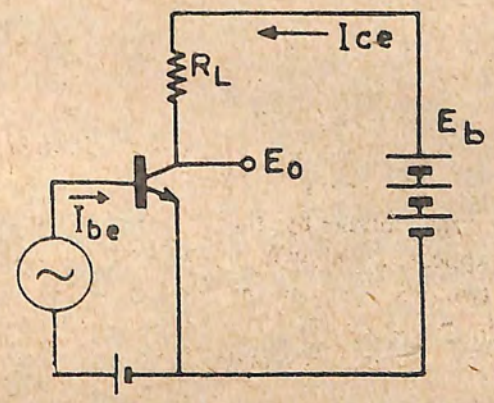


Fig. 20.20 Transistor amplifier.

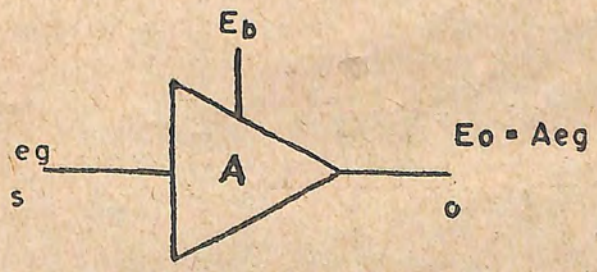


Fig. 20.21 Symbol for amplifier voltage amplification.

opera on. It, in fact, converts the steady energ of the battery into a varying one in accor ance with the signal.

A close look at the circuit in Fig. 20.19(a) show that as the grid potential rises towards a positive value, the current I_a will tend to increase, resulting in an increase in the potential drop across R_L. Since the sum of the potential across R_L and that across the triode due to r_p must be constant and equal to the battery potential E_b, it is natural that the potential across the triode wall tend to reduce. Thus the output voltage will swing so as to lower its value when the signal has a tendency to rise. Let us assume that we have applied a sinusoidal control voltage e_g as shown in Fig. 20.22 (a) to the grid. Such a voltage will swing it in either direction. The output voltage by the above argument will appear as shown in Fig. 20.22 (b). The two voltages can be considered as shifted in phase by 180° In other words, there is a phase reversal between V_g and E_o

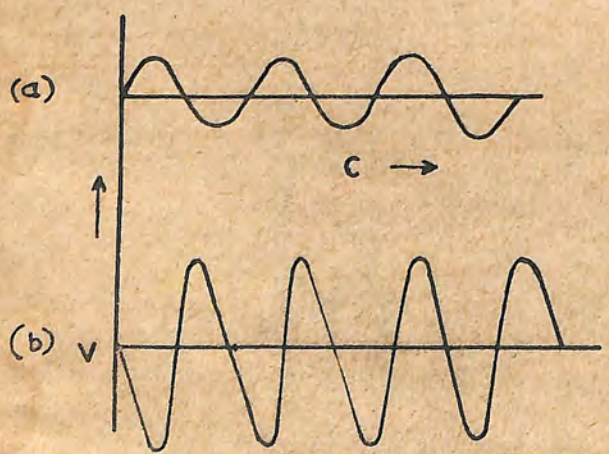


Fig. 20.22 (a) Input voltage (b) Output voltage.

EXAMPLE 20.3

(a) Find the gain of an amplifier using a triode for the valve of $R_L = 100$ k Ω. The triode has plate resistance $r_p = 15$ k Ω and has $\mu = 60$.

Solution

$$\text{Amplifier gain } A = \frac{\mu R_L}{r_p + R_L}$$

In the above example

$\mu = 60$, $R_L = 100$ k Ω; $R_p = 15$k Ω

$$A = \frac{60 \times 100 \times 10^3}{(15 \times 10^3) + (100 \times 10^3)}$$

$$= \frac{60 \times 100}{115} = 52.2$$

(b) *Oscillators*

Let us now consider a simple application of amplifiers by which oscillatery voltages are generated. Energy or voltage in oscillatory form have many applications. For example, radio waves which are used as carrier in radio communication are electromagnetic, oscillatory waves.

It is a general observation that the amplitude of oscillation of either a pendulum or that of an oscillating loaded spring tends to reduce gradually with every oscillation. This damping of the oscillations depends on the various ways in which the energy is lost. In case of the pendulum, for example, the losses would be due to friction between the air and the oscillating bob. The loss of energy can also be internal as in the case of oscillating spring.

An equivalent of the mechanical oscillator is an electrical circuit called tank circuit and consists of an inductance 'L' and a capacitance 'C' connected in parallel. An electrical energy put into it alternately oscillates between the magnetic energy in the inductance and electrostatic energy in the

capacitor. This is similar to the periodic conversion of the kinetic energy of the pendulum to the potential energy and vice versa. As in the case of losses of energy in a pendulum, the electrical energy stored in the tank circuit is gradually lost to the internal resistance of the inductance, etc. Thus we have an exact analogue of mechanical oscillators in electrical circuits. The frequency of oscillations of the circuit is given by

$$f = \frac{1}{2\pi\sqrt{LC}} \qquad ...(20.7)$$

For applications such as radio transmitter, we need to maintain oscillatory energy of constant amplitude for long time. It is therefore necessary, if we were to use say the tank circuit to provide oscillatory voltage, that some arrangement must be made to make up for the exact amount of energy lost. This can be done efficiently only if the energy fed into the circuit for the purpose has the same frequency as the frequency of oscillation and the way of feeding should be such as to aid the existing oscillation. The oscillations would otherwise die faster if these conditions are not adhered to.

Amplifiers essentially amplify small signal voltages and convert d.c. supply energy into a large signal energy. They can be effectively used to maintain the constant level of oscillations if properly arranged to satisfy the conditions given above. An oscillator using amplifiers is shown in Fig. 20.23.

The amplifier 1 amplifies the oscillating voltage across the tank circuit and inverts it through 180° as explained earlier. The second amplifier amplifies it further and inverts the voltage again thus bringing it in phase with the oscillatory voltage across the tank circuit. We thus have higher energy of

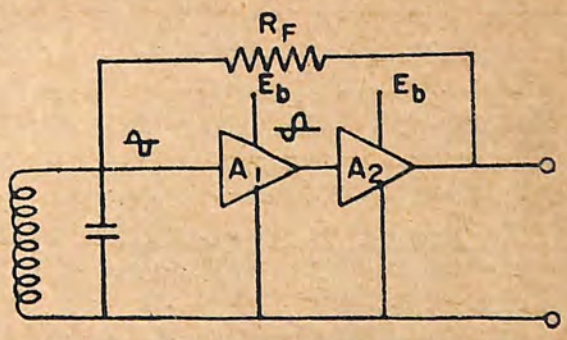


Fig. 20.23 Oscillator with amplifier.

the appropriate frequency in a phase to assist the oscillations already existing in the tank circuit. A small but adequate amount of this output energy is fed to the tank circuit. The amount of energy fed back from the output to the tank circuit thus is controlled by R_F.

The energy for oscillations is no doubt derived from the d.c. supply used for operating the amplifiers.

20.9 Manipulation of Position : TV and CRO Tubes

We have so far considered the devices which use manipulation of number of electrons or charge carriers to produce desired effects, for example, the amplification. It was already mentioned that electrons can also be manipulated in such a way that their positions can be changed. This type of positional manipulation is effectively used in television and cathode ray oscilloscope tubes which are devices capable of presenting visible pictures.

The main function of these tubes is to transform electrical signals, available in the form of varying voltages or currents, into a visible form. For example, a picture on a television is composed of strips of electrical signals systematically arranged across the face of the tube.

It is necessary in that case, that the beam of electrons which is used to draw such a picture should be available in a fine point similar to the tip of an artist's painting brush. This picture, however, must be made visible for the benefit of the observer. This is generally done by using a layer of materials called *phosphors* on the inside of the face of the tube. The phosphors absorb the kinetic energy of the electrons striking them and emit light proportional to the number of electrons and their kinetic energy. In the case of the television tube, a typical phosphor emits nearly the white light. The phosphors also have to make sure that the picture drawn earlier on the face should vanish before the next is drawn. This is particularly essential in the case of television picture which continually change due to the movements of actors and backgrounds. The cathode ray oscilloscope tube on the other hand is used for tracing waveform produced by voltages which vary in strength and time. The phosphors used for these tubes generally produce pleasant green colour. For some applications, orange colour is prefered.

Fig. 20.24 shows an arrangement of electrodes used in a oscilloscope tube. In this device, as in the case of a triode, we have a thermionic source of electrons, k, and a grid

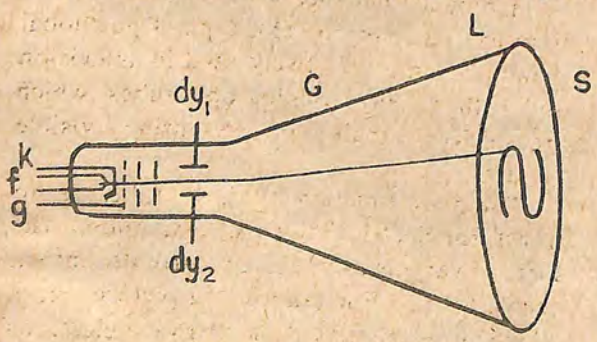


Fig. 20.24 Oscilloscope tube
g—grids, k—cathode f—filament,
G—glass bulb, L—thin layer of phosphor,
S—screen, dy_1, dy_2 —parallel plates.

g in front to control the number of electrons striking the screen on application of a negative potential with respect to the cathode. Similar to the cylindrical anode of the triode, whose job is to collect every electron that strikes it, this tube also has an anode, a_1, called accelerating anode, which has a small aperture in it. Some of the electron accelerated towards it due to the high positive potential pass through this aperture, acquiring velocities depending upon the accelerating potential. These electrons, obviously have a tendency to diverge from each other and as stated earlier must necessarily be focussed to a point. A second anode called focussing anode, a_2 is placed after a_1 and an appropriate potential is applied to it. The focussed beam of electrons thereafter passes between a pair of plates dy_1 and dy_2 and strikes the screen to produce a fine spot on it. In the absence of any potential applied between dy_1 and dy_2, the beam will travel in a straight line and the spot will be nearly in the centre of the screen. If on the other hand a d.c. potential were applied to these plates the beam will shift upward or downward depending upon the direction of the electric forces due to the applied potential. It is easy to imagine that to reach any point on the screen with the beam, it is necessary to have another pair of plates perpendicular to the former pair. The pair of plates which deflect the beam horizontally are called x deflection plates, and those which deflect vertically are called y deflecting plates.

In order to use such a tube, it is necessary that proper voltages are applied to the x and y deflection plates. For tracing waveforms, which are essentially voltages varying in strength and time, it is customary to apply to the x plates a voltage which would sweep the beam across the face of the screen in such a way that the beam traverses it at a

constant velocity. This type of sweep is called a linear sweep. It is, however, necessary to bring the beam back to its original position so that the waveform can be repeatedly traced on the screen for observation. This is achieved by applying a specially obtained voltage with a waveform similar to a sawtooth and is shown in Fig 20.25. The voltage to be observed is now applied to the y plate. Fig. 20.24 shows a sinusoidally varying waveform appearing on the face of the screen when a linear sweep voltage is applied to the x plates and a sinusoidal voltage is applied to the y plates.

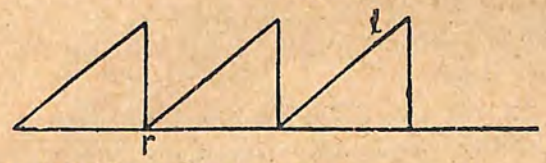


Fig. 20.25 Saw-teeth wave form,
r—return of a beam, t—linear sweep.

In the TV tube, we have exactly similar arrangement consisting of a cathode, a grid and accelerating and focussing anode. Here the deflection is achieved magnetically by a pair of coils kept outside on the neck of the tube. The arrangement is shown in Fig. 20.26. In this case both the deflection coils are used to sweep the beam across the face of the TV tube, in such a way that the beam traces from the left top corner and moves horizontally across to the other edge. On reaching that edge, it again jumps back to the left edge but starts a bit lower than the

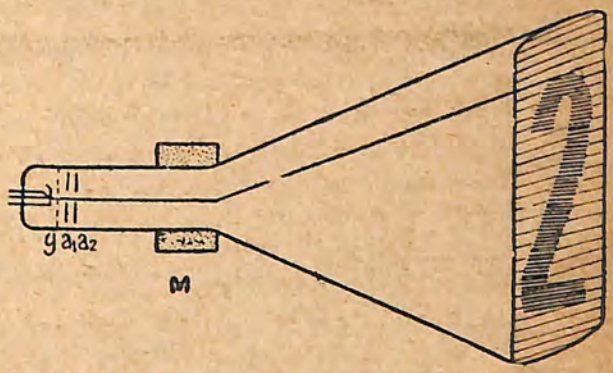


Fig. 20.26 TV tube
M—magnet, a_1, a_2 = accelerating and focussing anodes, g—grid.

previous line already drawn. The second line is again traced and the third and so on. This type of sweep across the whole face of the tube is known as raster. The picture that is obtained is produced by manipulating the intensity of the electron beam striking the screen. This will produce dark, bright and various shades of brightness on the screen. This device, in particular, is an example of using both the positional and number manipulation techniques.

In this unit, we have shown that electrons can be utilised in devices so as to fully exploit the speed with which they can be manipulated. We have also described a few applications of some of the simple devices, these are just a few examples of how electronics has helped to achieve the speeds of actions and revolutionise many fields of social importance.

Exercises

20.1 Find the values of r_p from the characteristic curve given in figure 20.14 for the operating conditions given below :

(a) $V_a = 170V$; $I_a = 8.5$ mA ; $V_g = -1V$ (10.5 k Ω)

(b) $V_a = 250V$; $I_a = 10$ mA ; $V_g = -2V$ (11k Ω)

20.2 Obtain the value of g_m from the characteristic curve given in figure 20.15.
(a) $V_a = 200V$; $I_a = 11.5$ mA ; $V_g = -1V$ (6.7 mS)
(b) $V_a = 250V$; $I_a = 10$ mA ; $V_g = -2V$ (5.5 mS)

20.3 Obtain the values of μ, g_m & r_p from the characteristics and verify the result $\mu = r_p \times g_m$ for the operating conditions given below :
(a) $V_a = 300V$; $I_a = 5$ mA ; $V_g = -4V$ (52.5, 5.1mS 17 kΩ)
(b) $V_a = 400V$; $I_a = 2.5$ mA ; $V_g = -6V$ (55, 2.4 mS 23 kΩ)

20.4 Find the value of β for the transistor (U BF115) from the characteristics given in figure (20.18) for,
(a) $V_{ce} = 1V$ and $I_b = 25\ \mu A$; 150 μ A; 200 μA (100)
(b) $V_{ce} = 1.5V$ and $I_b = 50\ \mu A$; 200 μ A; 400 μA (93)

20.5 Find the value of β for the transistor (BF115) from the characteristics given in figure (20.18) for,
(a) $V_{ce} = 2V$ and $I_b = 25\ \mu A$, 100 μA; 250 μA
(b) $V_{ce} = 2.5V$ and $I_b = 100\ \mu A$; 200 μA; 300 μA
(c) $V_{ce} = 3V$ and $I_b = 25\ \mu A$; 50 μA; 75 μA
(d) $V_{ce} = 4V$ and $I_b = 25\ \mu A$; 200 μA; 400 μA (100, 93, 100, 100)

20.6 The r_p of a triode is 15 kΩ and the gain A of an amplifier using the triode is 50. The load resistance is 100 kΩ. What is the μ of the triode ? (57.5)

20.7 Find the gain A of an amplifier. The amplifier makes use of the tube with $g_m = 5.5$ mS and $r_p = 11$k Ω for the following load resistances.
(a) $R_L = 10$ kΩ
(b) $R_L = 25$ kΩ
(c) $R_L = 100$ kΩ (28,8 40.75, 60.5)

20.8 Find the gain of the amplifier for the operating condition for the triode as given in example given above for the following values of R_L.
(a) $R_L = 100$ kΩ
(b) $R_L = 10$ kΩ
(c) $R_L = 100$ kΩ
(d) $R_L = 1$M kΩ

20.9 Find the gain of an amplifier using transistor (BF115) for $\beta = 98$;
Following are the various load resistances used and the internal resistance of the transistor is 0.6 kΩ
(a) $R_L = 10$ kΩ
(b) $K_L = 50$ kΩ
(c) $R_L = 1$ MΩ (94, 96, 98)

20.10 Find the gain of an amplifier using the characteristic of transistor (BF115) given in figure (20.18), for the condition $V_{ce} = 2V$ $I_b = 25\ \mu A$ & 100 μA. R_L the load resistance is equal to 100 kΩ. Internal resistance of the transistor is taken as 0.6 kΩ. (100, 73)

Appendix

Mathematical Notes

Differential Calculus : The concepts of velocity and acceleration have already been discussed in the previous class with the help of graphical methods. Let us extend the above knowledge in determination of velocity and acceleration under different situations.

Velocity : Let us consider a particle to be moving along positive x-direction as shown in Fig. 1. At a time t_1 the particle is at a point A and at a distance x_1 from the origin. Let the particle reach point B at a distance x_2 at the time t_2. Here we see that the distance travelled by the particle depends on the time. Mathematically we say that the distance travelled by the particle is a function of the time interval.

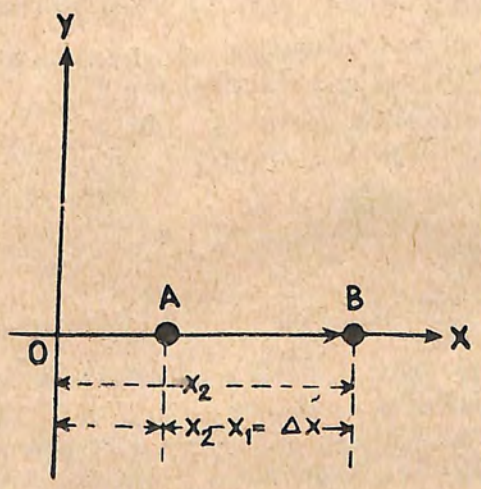


Fig. 1

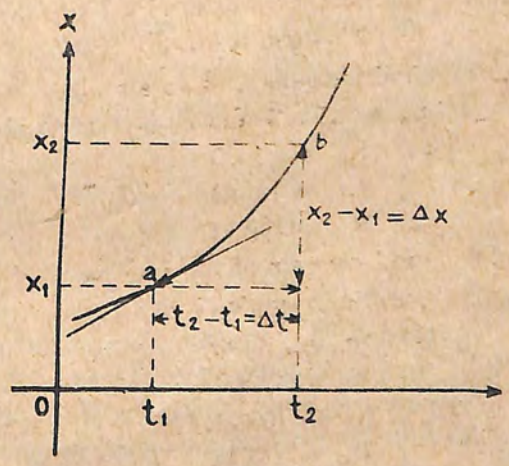


Fig. 2

If the distance travelled by the particle in the above case is plotted on a graph as a function of time we get the graphs as shown in the Fig. 2. From the graph the ratio of the particle displacement (which is a vector) $x_2 - x_1$ to the time interval $t_2 - t_1 = \Delta t$ gives the average velocity (also a vector) between points a and b.

$$\text{Average velocity} = \frac{\text{Displacement}}{\text{Time}}$$

$$v = \frac{x_2 - x_1}{t_2 - t_1}$$

$$v = \frac{\Delta x}{\Delta t}$$

As we see, the ratio $\frac{\Delta x}{\Delta t}$ is also the slope of the cord a b.

Let us consider a case in which velocity of the particle changes with time. The above method will only give the value of average velocity. The actual value of the velocity of the particle at any point of its path or at any instant of time is called its instantaneous velocity. Let us find the instantaneous velocity of the particle (Fig 1 and 2) at the point A. The average velocity between point A and B is associated with the entire displacement Δx and with the entire time interval Δt. Move the point b closer and closer to the point a (Fig. 2) and calculate the average velocity for shorter and shorter time intervals. When the time Δt is very short such that $\Delta t \to 0$, then points a and b are very close to each other and the line a,b nearly coincides with the tangent line at the point a. Then the velocity calculated by the ratio $\frac{\Delta x}{\Delta t}$ in the case when $\Delta t \to 0$ gives the instantaneous velocity and is written as $\frac{dx}{dt}$. Here $\frac{dx}{dt}$ is called the derivative of x with respect to t.

Mathematically we can express this statement as

$$v = \lim_{\Delta t \to 0} \left(\frac{\Delta x}{\Delta t} \right)$$

$$= \frac{dx}{dt}$$

In the Fig. 1 as the point B approaches the point A, point b approaches point a in the Fig. 2. In the case when $\Delta t \to 0$ (which is also khown as limit of Δt) the slope of chord ab equals the slope of the tangent to the curve at the point a. Consequently the intantaneous velocity at any point of x—t graph is equal to the slope of the tangent to the graph at that point.

Acceleration : When the velocity of the body changes continuously the body is said to move with accelerated motion. Fig. 3 shows a particle moving along the x-axis. The vector v_1 represents its instantaneous velocity at a point A and the vector v_2 represents its instantaneous velocity at point B. Fig. 4 is a graph of instantaneous velocity v as a function of time. The points a and b correspond to the points A and B of Fig. 3. The average acceleration of the particle as it moves from A to B is defined as the ratio of the change in instantaneous velocity to the elapsed time.

$$\text{Average acceleration} = \frac{\text{Change in instantaneous velocity}}{\text{elapsed time}}$$

$$a = \frac{v_2 - v_1}{t_2 - t_1} = \frac{\Delta v}{\Delta t}$$

The instantaneous acceleration of a body, that is, the acceleration at a given

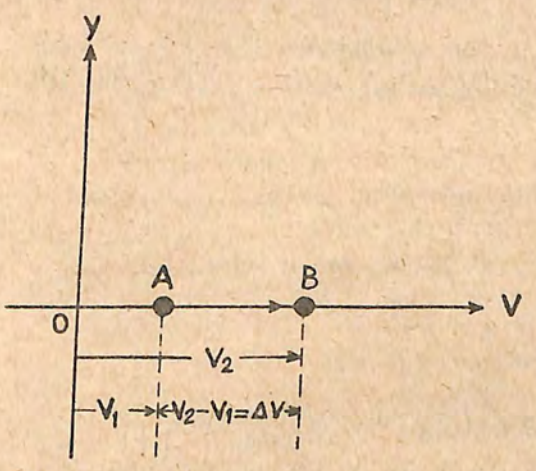


Fig. 3

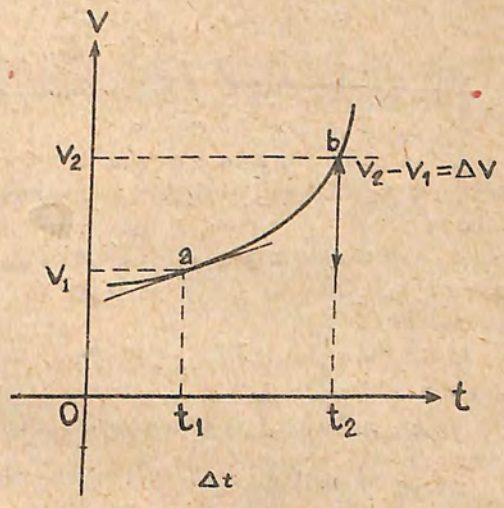


Fig. 4

instant of time, is defined in the same way as instantaneous velocity. The instantaneous acceleration at the point A is defined as the limiting value of the average acceleration when point B is taken very close to A.

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t}$$

$$= \frac{dv}{dt}$$

Also $a = \frac{d}{dt}\left(\frac{dx}{dt}\right)$ $\left(\text{because } v = \frac{dx}{dt}\right)$

$$a = \frac{d^2x}{dt^2}$$

Here $\frac{d^2x}{dt^2}$ is called second derivative of x with respect to t. Also the instantaneous acceleration at any point on v—t graph is equal to the slope of the tangent to the graph at that point.

From the above discussion we see the important use of derivatives in the study of Physics. In general any variable quantity y which depends for its value upon another variable x is said to be a function of x and is written mathematically as.

$$y = f(x)$$

Here f (x) does not mean f. x but it is the way of representing the function of x.

EXAMPLE

The area of a square denoted by y depends for its value upon another variable x which is the side of the square. In this case

$$y = x.x$$
$$= x^2$$
$$y = f(x) = x^2$$

Definition of a derivative of a function : Let y be a function of x, i.e., $y = f(x)$. x increase by $\triangle x$. Let the corresponding increase in y be $\triangle y$. Then the new Let value of function will be

$$y + \triangle y = f(x + \triangle x)$$
$$\therefore \triangle y = f(x + \triangle x) - f(x) \quad (\because y = f(x))$$
$$\frac{\triangle y}{\triangle x} = \frac{f(x + \triangle x) - f(x)}{\triangle x}$$

In the limit $\triangle x \to 0$

$$\frac{dy}{dx} = \underset{\triangle x \to 0}{\text{Lim}} \left(\frac{\triangle y}{\triangle x} \right)$$

Hence derivative of a function y is defined as the limit of $\frac{\triangle y}{\triangle x}$ when $\triangle x \to 0$

Some of the important derivatives which are of use in the present course are listed below. The teacher may derive them in the class.

(1) $\frac{dc}{dx} = 0$ (where c is a constant quantity)

(2) $\frac{d}{dx}(x) = 1$ $\left(\text{we write } \frac{dx}{dx} \text{ as } \frac{d}{dx}(x) \right)$

(3) $\frac{d}{dx}(x^2) = 2x$

(4) $\frac{d}{dx}(cx^2) = c \frac{d}{dx}(x^2) = 2c.x$

(5) $\frac{d}{dx}(u + v) = \frac{du}{dx} + \frac{dv}{dx}$ (u and v are functions of x)

(6) $\frac{d}{dx}(u - v) = \frac{du}{dx} - \frac{dv}{dx}$

(7) $\frac{d}{dx} \sin x = \cos x$

(8) $\frac{d}{dx}(\cos x) = -\sin x$

(9) $\frac{d}{dx} \sin(\omega x + c) = \omega . \cos(\omega x + c)$

(10) $\frac{d}{dx} \cos(\omega x + c) = -\omega . \sin(\omega x + c)$

EXAMPLE

For a particular motion of a particle, the distance covered is,

$$s = kt^2$$

Where k is a constant and t is the time interval. To calculate velocity we find derivative of s with respect to t.

$$v = \frac{ds}{dt}$$
$$= \frac{d}{dt}(s)$$
$$= \frac{d}{dt}(kt^2)$$
$$= k.\frac{d}{dt}(t^2)$$
$$= k.\ 2t = 2kt$$

We have seen that if the coordinate x of a particle moving on the x-axis is given as a function of time, the velocity can be found by taking the derivative of x with respect to t, i.e., differentiation, In other words

$$v = \frac{dx}{dt}$$

A second differentiation gives the acceleration a

$$a = \frac{dv}{dt}$$

Integral Calculus : We now consider the process of finding the coordinate x as a function of t if the velocity is given or of finding v as a function of t if acceleration is given. This can be done by the method of integral calculus, which is the mathematical inverse of differentiation.

If acceleration a (t) as a function of time is given then we know

$$\frac{dv}{dt} = a(t)$$

Here a (t) is not a multiplied by t but denotes that a is a function of t. Taking the denominator dt of the differential on the right hand side,

$$dv = a(t)\,dt$$

In order to find v as a function of t, we integrate the differential expression. The integration operation is indicated by writing the integral sign $\int$ in front of the given differential expression.

$$\int dv = \int a(t)\,dt$$

or

$$v = \int a(t)\,dt + C$$

Where C is any arbitrary constant called the constant of integration whose value can be determined if the velocity is known at any time.
If at time $t = 0$ the particle velocity $v = v_1$ then

$$C = v_1$$

Let us assume that the acceleration does not depend upon time. Then on evaluating the above integral, the velocity v (t) as a function of t can be found, i.e.

$$v = at + v_1$$

Since $\frac{dx}{dt} = v(t)$

we also have $dx = v(t)\,dt$

$$\int dx = \int v(t)\,dt$$

or

$$x = \int v(t)\,dt + C$$

the value of integration constant can be determined if the coordinate of the moving particle is known at any time.

If the acceleration is given as a function of x we can write the expression for the acceleration in the following ways.

$$a(x) = \frac{dv}{dt}$$

$$= \frac{dv}{dt} \cdot \frac{dx}{dx}$$

which may be written as

$$a(x) = \frac{dx}{dt} \cdot \frac{dv}{dx}$$

but $\frac{dx}{dt} = v$

$$\therefore \quad a(x) = v \cdot \frac{dv}{dx}$$

or $$v \cdot \frac{dv}{dx} = a(x)$$

$$v dv = a(x)\, dx$$

On integrating, this becomes

$$\frac{v^2}{2} = \int a(x)\, dx + C$$

In order to evaluate the integral of a given differential expression it is useful to remember some integration formulae. Some of the formulae which are of use in the study of Physics in the present course are listed below :

1. $\int dx = x + c$
2. $\int a dx = ax + c$ (a is **not** a function of x)
3. $\int x dx = \frac{x^2}{2} + c$
4. $\int \sin x\, dx = -\cos x + c$
5. $\int \cos x\, dx = \sin x + c$

Integration is basically addition. Graphically an integration can be considered as being the addition of small areas under a curve and the integral can be interpreted as an area. Let the area under the velocity-time graph in Fig. 5 between the vertical lines at t_1 and t_2 be subdivided into narrow rectangular strips of width $\triangle t$. The ordinate of the graph at any time t equals the instantaneous velocity v at that time. If the velocity remained constant at this value, the displacement $\triangle x$ in the time interval between t and $t + \triangle t$ would equal $v \triangle t$. But this is the area of the shaded strip, since its height is v and its width is $\triangle t$. The sum of the areas of all such rectangles, between t_1 and t_2 is approximately equal to the total displacement $x_2 - x_1$ in this time interval.

$$x_2 - x_1 = \Sigma v \triangle t$$

Here the sign Σ denotes summation over all the rectangles. The smaller the time intervals $\triangle t$, the more closely does $v \triangle t$ approach the actual displacement. In the limit, as $\triangle t$ approaches zero, the sum of the areas become exactly equal to the total area under the curve, and to the total displacement $x_2 - x_1$. The limit of the sum of the areas is called the definite integral from t_1 and t_2. So

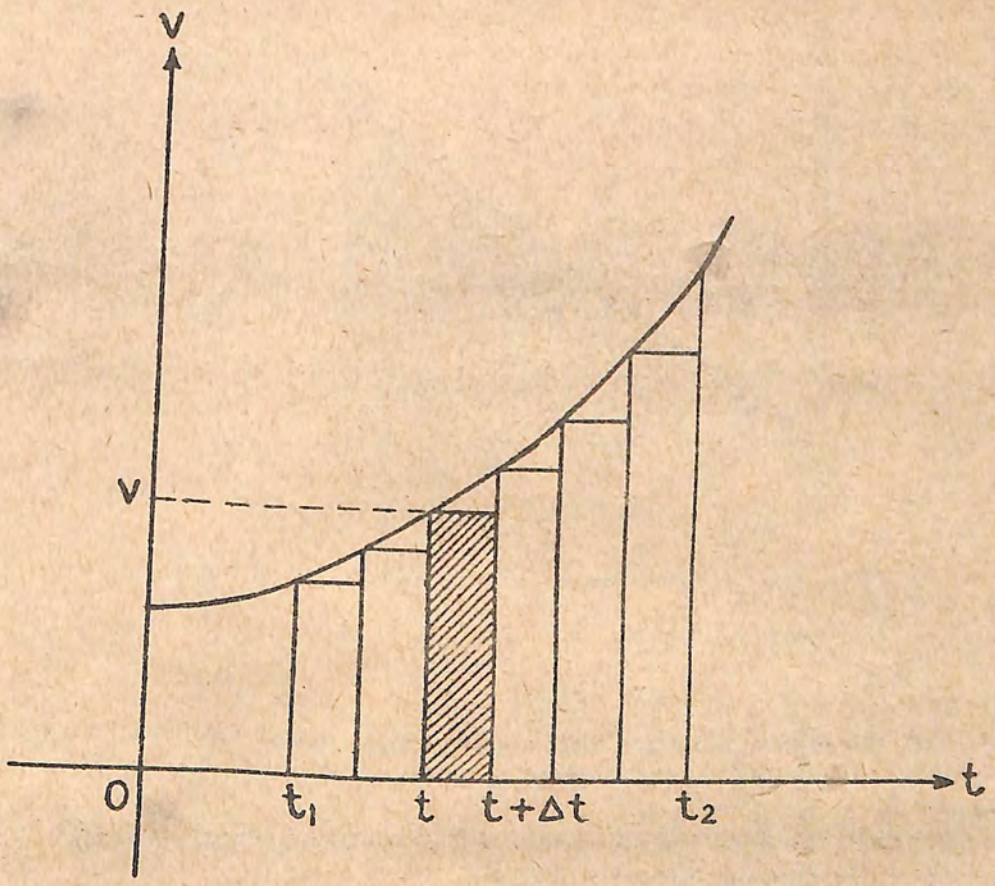


Fig. 5

$$x_2 - x_1 = \int_{t_1}^{t_2} v\,dt$$

$$= \left[vt + c \right]_{t_1}^{t_2} \quad (\text{if } v \text{ is constant})$$

$$= (vt_2 + c) - (vt_1 + c)$$

$$= vt_2 - vt_1$$

In this case when the limits of integration are given, the constant of integration disappears.

From the above discussion, we see that the displacement in any time interval is equal to the area between velocity-time graph and time axis, bounded by vertical lines at the beginning and end of the interval.

Similarly, the area under an acceleration-time graph can be used to show that the change in velocity in any time interval is equal to the area between the acceleration and time graph and the time axis bounded by vertical lines at the beginning and end of the interval.

EXAMPLE 3

Use the definite integral to find at any time velocity and the coordinate

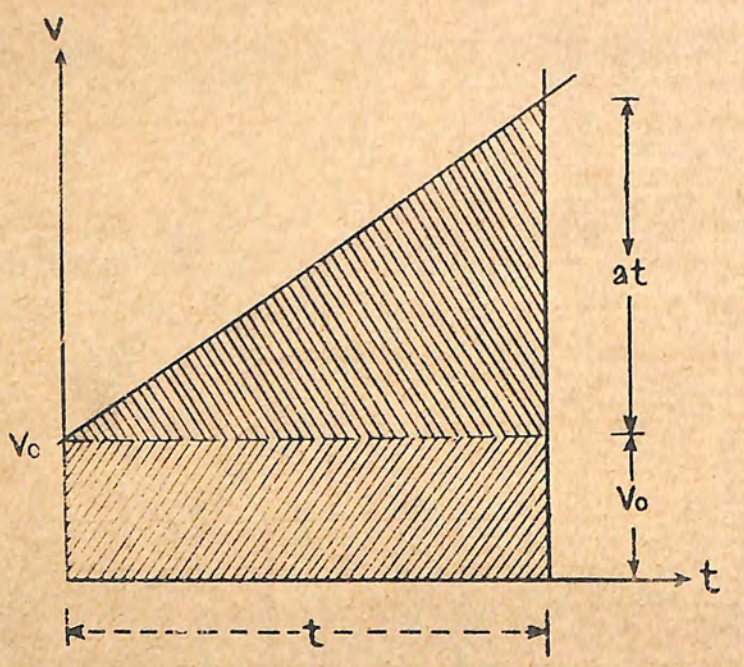


Fig. 6

of a body moving on the x-axis with constant acceleration. The initial value of velocity is v_0 and the initial coordinate is zero.

Solution :

The limits of integration are $t_1=0$, $v_1=v_0$, $x_1=0$ and $t_2=t$, $v_2=v$, $x_2=x$

Integrating the expression between the limits

$$dv=adt$$

$$\int_{v_1}^{v_2} dv = \int_{t_1}^{t_2} adt$$

$$v_2-v_1 = \int_0^t adt = at$$

or $v-v_0=at$

$$\text{Also } \int_{x_1}^{x_2} dx = \int_{t_1}^{t_2} v\,dt$$

$$x-x_1 = \int_0^t (v_0+at)\,dt$$

$$x=v_0t+\tfrac{1}{2}at^2$$

It is not always necessary to use the method of integral calculus to find the area under a graph. Fig. 6 is the velocity-time graph for motion with constant acceleration. The area under the graph, between $t=0$ and $t=t$, can be subdivided into a rectangle and a triangle. The area of rectangle is v_0t and that of the triangle is $\frac{1}{2}t\times at=\frac{1}{2}at^2$. Since the displacement equals the total area

$$x-x_0=v_0t+\tfrac{1}{2}at^2$$

[illegible]

एकलव्य
अध्ययन केन्द्र मालाखेड़ी
होशंगाबाद फोन- 277267

Eklavya Library
Class No. 2133
Book No.
Shelf No. 64.5
Acc. No. 13,373

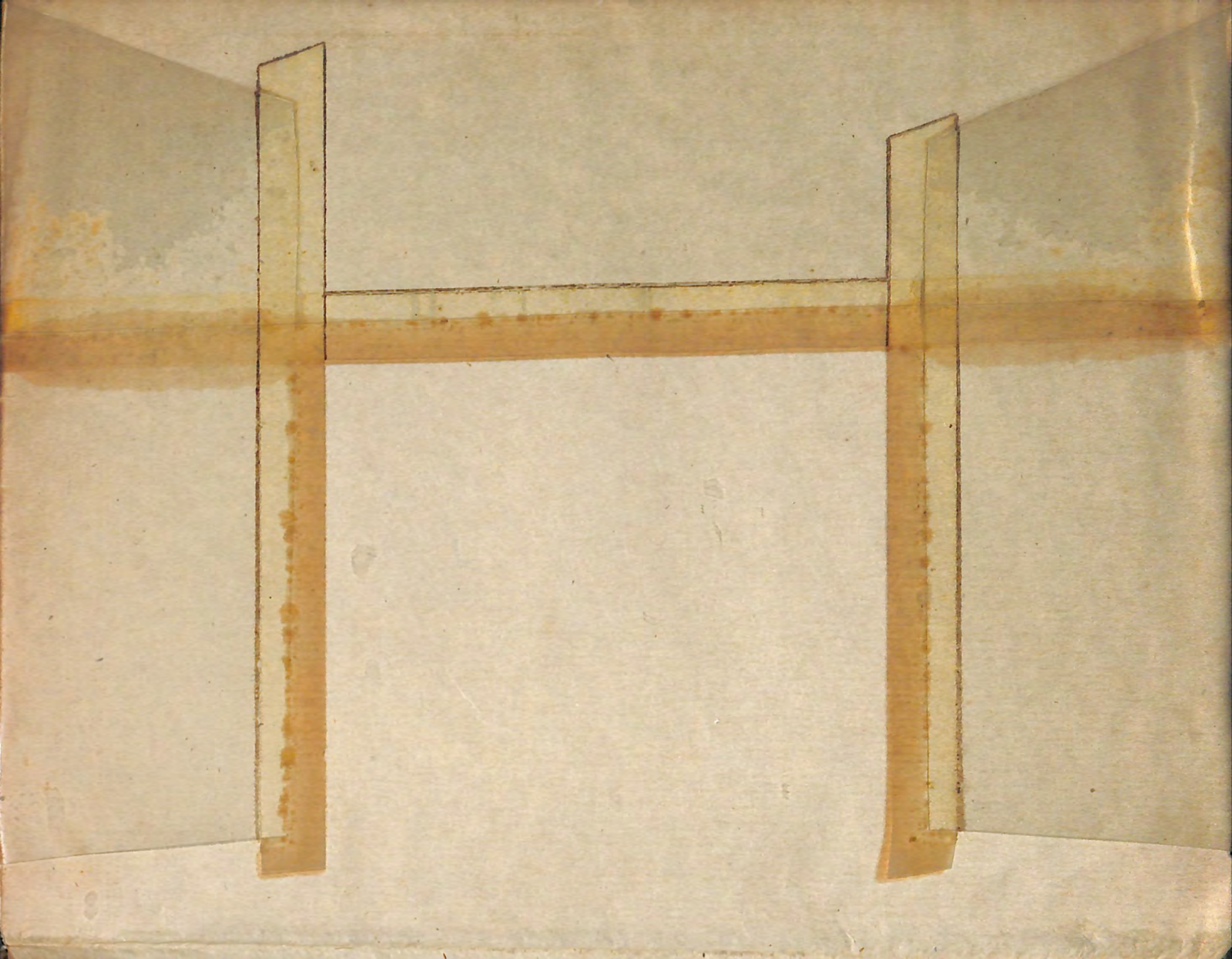
EKLAVYA LIBRARY, HOSHANGABAD

Ph. 277267, 277268

email : hoshangabad.library@eklavya.in

DUE DATE	DUE DATE

निर्धारित दिनाँक तक पुस्तक जमा न करने पर विलम्ब शुल्क देय होगा।



राष्ट्रीय शैक्षिक अनुसंधान और प्रशिक्षण परिषद्
NATIONAL COUNCIL OF EDUCATIONAL RESEARCH AND TRAINING